

Surrender on technology support

The Clinton administration is sticking to its desire to fund key technology developments in industry, in the teeth of political and scientific opposition, and with little past success to build on. The president should give way.

THE relationship between science and technology tightens all the time and in many vital disciplines the distinction has almost disappeared. But in political circles there is a significant difference: government funding for science has universal support (at least in principle), but funding for technology does not. Hence the bitter battle in the United States between the Republican Congress and President Bill Clinton over the Advanced Technology Program (ATP), a scheme for giving grants to private companies to share the cost of developing technologies (not products) in key areas.

The ATP was born as a small, experimental programme under the Republican administration of President George Bush, but was targeted for rapid expansion by Clinton on his election in 1992. Its budget and scope then grew in leaps and bounds, from \$68 million in 1993 to \$410 million last year, and was to have reached \$750 million in 1997. Clinton and the Congress are now locked in battle about whether to spend \$300 million or zero on the ATP this year.

The National Institute of Standards and Technology (NIST), which administers the programme, has worked with industry to establish distinct programme areas for the expanded ATP. Within these programmes, companies compete for grants worth \$2 million or more for individual projects. This process has been widely praised. The recent suggestion from Robert Walker, chair of the Science committee of the House of Representatives and a leading opponent of ATP, that its grants were being distributed to particular companies as political patronage, was an unsubstantiated slur on NIST staff that adds nothing to this debate.

But more serious opposition to the programme has come from the scientific community, parts of which feel (wrongly) that money cut from ATP will flow to science. Last November, a National Academy of Sciences panel chaired by Frank Press, formerly science adviser to President Jimmy Carter, said it was "sceptical" whether programmes such as ATP were "the most efficient use of scarce federal R&D dollars", and that such programmes should continue "only if the case is convincingly made that the government is the funder of last resort for an important enabling technology and ... only on an experimental basis".

Under fire from George Brown (Democrat, California), the senior Democrat on the House Science committee, Press backtracked somewhat from this position, saying that he wanted only more "evaluation" of ATP (see *Nature* 380, 5; 1996). He should have stuck to his guns. At funding of hundreds of millions of dollars a year, ATP is not an experiment: it is an expensive act of faith. Clinton administration officials have argued that the programme needs a critical mass of \$1 billion a year (one per cent of US industry's total R&D investment) to have a measurable impact on US industrial competitiveness. But no-one will ever know if they are right: there is no accepted design for the "experiment", no agreed end-points, and no control. All we have is experience — and the history of government support for technology development is patchy. The less impressive examples include many tactical efforts to pass grants to industrial companies to help them do what they ought to be doing for themselves. In the United States, where neither private capital nor entrepreneurial spirit is in short supply, such efforts are harder to justify than elsewhere.

At a time when the US government has no money for the excellent university science funded by the National Science Foundation

(see page 187), no money to restore a tattered social fabric, and less than nothing in the bank, the onus remains very firmly on Clinton to prove that the Advanced Technology Program deserves the priority he has given it. That will be a tough task. □

German consensus undermined

Pressures following reunification threaten a fine example of collaborative support for science.

THE long-established system of support for research in Germany, involving both the federal and the 16 *Länder* governments, is looking increasingly fragile. The most recent symptom of decline arose earlier this month when Edmund Stoiber, the powerful prime minister of Bavaria, threatened to pull out of the finely balanced arrangement designed, as a product of Germany's post-war constitution, to avoid undue concentrations of power.

Within this system the federal and *Länder* governments share, according to fixed formulae, the costs of the four so-called 'pillars' of German research: the Max Planck Society (basic research), the Fraunhofer Society (applied research), the national research centres and the so-called Blue List institutes. Sharing costs means sharing political influence. It gives the federal government a fair say in steering national research and allows it to ensure that effort is well distributed throughout the country. In the case of the Max Planck Society, the host *Land* pays only a quarter, rather than a half, the rest being provided from a common fund, to which all *Länder* must contribute, but over which the society has full control. This system weakens the political influence of the host *Land*, giving the society the freedom to decide when and where to place, or close, research institutes on scientific grounds alone.

Until recently, complaints from governments of *Länder* with few Max Planck institutes, vexed at having to subsidize institutes in better-endowed *Länder* via the common fund, have been muffled. But the financial pressures of reunification, and the need to establish Max Planck institutes in the five new, and poor, *Länder*, are changing all that. A majority vote at the last meeting of the *Länder* prime ministers supported a move to abandon the common fund altogether, and make each host *Land* pay its full 50 per cent quota.

Stoiber, who likes to regard prosperous Bavaria as the country's science capital, was right to resist that proposal — although he did so for the wrong reason: he appears to fear only the steep bill that Bavaria would face if it lost its common fund. More questionable is his threat to pull out of the joint funding system altogether — a move that could cause collapse in 1999, when the system as a whole will be up for renewal.

The federal government is expected to form its views on the common fund in June. It should join Stoiber in resisting the fund's abandonment. The Max Planck Society is sensitive to the needs of the *Länder*, but in establishing new institutes must be guided by local scientific strengths, sometimes in the face of political pressures. The common fund strengthens its hand in this respect. It represents a principle that is all too easily abandoned when times are hard, and much more difficult to reinstate later. □



1997 budget confirms US science faces a period of flat funding

Washington. The fourth and final budget of President Bill Clinton's first term was submitted on Tuesday to Congress. The budget, for the 1997 financial year which starts 1 October, promises increases of close to the rate of inflation (3 per cent) for the main science funding agencies, and the restoration of technology and environmental research programmes cut by Congress last year.

The National Institutes of Health (NIH) asks for a 3.9 per cent increase (see below), although two-thirds of this will be consumed by the construction costs of a new clinical centre building at the NIH campus in Bethesda, Maryland. The Department of Energy requests a 2.5 per cent increase in its non-weapons research budget, while the National Aeronautics and Space Administration (NASA), is asking for roughly the same as the current year (see page 188).

Most other research-funding agencies are requesting little more for 1997 than they received in their last properly-completed appropriation in 1995. This bears out fears that US science is facing a period of flat funding, steadily eroded by inflation.

The National Science Foundation (NSF),

for example, is requesting \$3.325 billion, only 3 per cent more than it received in 1995. The National Oceanic and Atmospheric Agency (NOAA)'s request for \$2.1 billion reflects the same trend.

The 1997 budget has been released even though the final budget figures for 1996 remain incomplete. Agencies which have an agreed 1996 budget, such as NIH, were able this week to present their proposed 1997 figures relative to that baseline. But other agencies could only compare their proposed budgets to an estimate of what they will receive this year.

In most cases, that estimate was based on the temporary funding they currently receive, which expires tomorrow (22 March). If Congress and the President cannot reach agreement on the 1996 budget, that temporary funding may be extended to 30 September, when the financial year ends.

The Department of Energy (DOE), whose research budget was cut sharply this year from \$2.7 billion to \$2.5 billion, is asking for an increase of only 2.4 per cent in its energy research programmes next year, and 2.7 per cent in its physics programmes.

Within these totals, the DOE wants to restore some money to its fusion energy programme, raising the latter's budget by 12 per cent to \$256 million. But it also plans to cut biological and environmental research by 7 per cent, to \$380 million.

The National Science Foundation (NSF) is emphasizing research spending over the educational and other activities in its budget request, which is up 4.6 per cent on the agency's estimated 1996 spending level of \$3.18 billion. NSF's support for research and related activities will grow by 8.7 per cent from the 1996 estimate, to \$2.47 billion.

Of the NSF's seven directorates, engineering does best, with a proposed 12 per cent increase, mostly for grants to independent investigators. But the NSF wants to eliminate funding for equipment and improvements at university laboratories, which Congress traditionally favours, and which received \$100 million in 1996.

The Clinton administration also promised to restore environmental and technology programmes that have been cut this year at the Department of Commerce. Ron Brown, the commerce secretary, concedes that his original objective of expanding the National Institute of Standards and Technology (NIST) to a \$1.4-billion agency this year had been "significantly revised". But he is pledging that he will continue with NIST's technology programmes.

NIST is seeking a budget of \$823 million in 1997, up sharply from the \$580 million which the agency will get this year at current spending levels. The figure would include \$345 million for the Advanced Technology Programme, which Congress wants to eliminate, and \$105 million for major laboratory construction work at NIST's site in Gaithersburg, Maryland.

The National Oceanic and Atmospheric Administration (NOAA) is requesting an 8.3 per cent increase in funding, from \$1.938 billion to \$2.099 billion. James Baker, NOAA's administrator, says he is happy with the request. But it concedes a lot of ground to the agency's critics in Congress: the Office of Oceanic and Atmospheric Research, which supports most basic research at NOAA, is demanding just \$233 million — sharply down from the \$271 million it requested last year.

The Department of Defense did not release details of its research budget plans this week. But its total spending on research, development, test and evaluation will be \$34.7 billion, down slightly from last year's \$34.9 billion.

Colin Macilwain

A shot in the arm for clinical research

Washington. President Bill Clinton has asked the US Congress to approve a 3.9 per cent increase in funding for the National Institutes of Health (NIH) in the \$1.64-trillion 1997 budget that he sent to Capitol Hill earlier this week.

Almost exactly two-thirds of the \$467-million increase, which raises the NIH's budget to \$12.4 billion, would go towards building a new, 250-bed clinical research facility on the NIH campus. The \$310 million for the proposed 250,000-square foot Clinical Research Center (CRC) would be spent over a period lasting until the centre's projected completion early next century.

Most of the remainder of the increase sought for the institutes — about \$135 million — will be used to boost the funds available for research grant applications received from non-NIH scientists. The NIH will fund 207 new multi-year grants in 1997, reaching a total of 6,827.

Meanwhile, routine inflationary increases for external scientists who already receive multi-year grants will be cut from 4 to 2 per cent. Harold Varmus, the director of NIH, calls this a "measurable

sacrifice" that researchers are likely to be willing to accept. And almost all other NIH programmes will see their funding either remain constant or decrease.

Varmus said on Monday that the construction of the new Clinical Research Center was essential. "This is the best year to do it. But we're doing everything we can to protect the extramural grant community," he said, pointing out that most of the increased funding is needed to replace a decrepit 1953 building that he recently described as an imminent "threat" to NIH patients and personnel. "We're getting that much because the president and the administration have decided it's important" to build the new centre, says Varmus.

Ralph Bradshaw, the president of the Federation of American Societies for Experimental Biology, which last month called for a 6.5 per cent increase in the NIH budget, said that although his organization felt that an increase of 6.5 per cent would have been the "appropriate number", any proposed increase at the NIH would obviously be viewed favourably.

Meredith Wadman

US budget envisages level funds for NASA, more for environment

Washington. The US National Aeronautics and Space Administration (NASA) is requesting roughly the same amount in next year's budget that it expects to receive from Congress this year, namely \$13.8 billion. This is 4.2 per cent less than the budget it received in 1995. It also marks a \$100-million decline from projections that were being made by the White House for NASA's 1997 budget as recently as last summer.

Spending on space science, an account that includes astrophysics, planetary exploration and space physics, would drop by 8.4 per cent from the expected 1996 appropriation, to a total of \$1.86 billion. This is largely due to a decrease in the funds required for the Cassini mission to Saturn and for the Advanced X-Ray Astrophysics Facility, both of which are nearing their launch dates. At \$498 million, the budget for life and micro-gravity sciences is slightly higher than this year, signalling a gradual increase in efforts aimed at developing experiments to be conducted on the space station.

NASA's request for the Mission to Planet Earth is \$1.4 billion, an increase of 8 per cent over the expected 1996 level, and an indication of the administration's determination to continue this embattled programme, which has recently come under some critical scrutiny in Congress. Funding for human space flight, which includes both the space station and the space shuttle, remains about level at \$5.86 billion.

In the Department of Interior, the administration also is holding close to budget lines drawn in Congress over the past several months. The 1997 request for the US Geological Survey (USGS) is \$746 million, including \$145 million for "natural resources research", the remnants of the now-defunct National Biological Service. The budget represents a 2.1 per cent increase over the USGS's expected 1996 appropriation.

The Fish and Wildlife Service is requesting \$83 million for work on endangered species. This includes \$7.5 million for listing new species, money that Congress virtually deleted for 1996.

The Environmental Protection Agency is seeking \$579 million for its research and development efforts, amounting to 9.8 per cent more than this year's expected Congressional appropriation, but only 84.8 per cent of the figure that the administration had originally requested for 1996. But the final EPA figure for this year was still up in the air at the beginning of this week; the White House has made environmental funding a key sticking point in its ongoing budget negotiations with Congress.

Tony Reichhardt

Ministers seek legal shelter from Canada's HIV inquiry

Montreal. Provincial governments in Canada are alleged to have demanded that Justice Horace Krever, who is heading a commission of inquiry into the distribution of HIV-infected blood in the 1980s, should sign a statement promising not to make findings of misconduct against their ministers or their predecessors.

When this request was denied, according to an affidavit filed by Marlyns Edwardh, counsel for the commission, the ministers are said to have ordered the lawyer to say they had not been adequately represented by legal counsel at the hearings.

The allegations come after provincial governments, the Canadian Red Cross Society, pharmaceutical companies and a host of individuals — including 34 former provincial ministers of health — have already taken legal action to prevent Krever from including information in his final report that could expose them to criminal or civil proceedings (see *Nature* 379, 479; 1996).

It now appears that they are not stopping there. Edwardh states in her affidavit that, in a series of telephone calls, William Craik, a lawyer representing 10 provinces and territories, said that the provinces were "exceedingly upset" that their ministers had been identified by Krever as being among those who might be linked to allegations of potential wrongdoing.

The affidavit states that, during the

course of this conversation, Craik "asked whether the commissioner would sign a statement indicating there would be no findings of civil or criminal responsibility against the ministers". Edwardh added that she recalled "a second request which related to the promise of not naming the ministers in the report".

Edwardh rejected the suggestion, and two days later, the commission received a letter from Craik refusing to accept the notices about possible wrongdoing on the grounds that he represented only the provinces, and not the ministers or government employees personally. According to Edwardh, all contact with individuals was, until then, required to be made through counsel appointed by the provinces, the federal government and the Canadian Red Cross Society.

Meanwhile, the Red Cross, which is also under fire from Krever, has been trying to force the commission to release some of its internal correspondence. The society wanted a list of all written material reviewed by Krever before sending out notices of possible misconduct, and related documents.

But a ruling released by Justice George Richards says there was no reason for such documents to be made public, as there is no indication that they contain new evidence. The laws of evidence, said Richards, are "not intended to authorize a fishing expedition".

David Spurgeon

Cash crisis halts nuclear power in India

New Delhi. A shortage of funds and a lack of political commitment have brought India's nuclear power programme to a halt, and a company set up by the government ten years ago to build and operate nuclear power stations has admitted that it is on the verge of liquidation.

A parliamentary committee that had been listening to the woes of nuclear industry for more than six months concluded in its report, published two weeks ago, that the government has been guilty of neglecting nuclear power. It has called for urgent action to bail the Nuclear Power Corporation (NPC) out of its financial crisis.

When it was set up in 1986, NPC was expected to build 22 nuclear plants, with a total installed capacity of 10,000 MW, by the year 2000. In the event, the government curtailed the programme, approving only four plants. This *ad hoc* policy, according to the committee, has put a heavy burden on the company, which had borrowed US\$750 million in the hope of obtaining a matching grants from the government.

NPC now owes \$100 million in interest

alone, and must find \$350 million to cover equipment orders placed for power plants that will not now be built. Another \$450-million-worth of nuclear components being held in storage will have to be sold as scrap. NPC says that, without government funding, it doubts whether it will be able to complete the four nuclear plants that are under construction.

India has estimated reserves of 75,000 tonnes of uranium and 360,000 tonnes of thorium, an element that can be converted into fissile uranium-233 on bombardment with neutrons. DAE had always argued that the best way to use these resources was through a three-tier strategy: burning uranium in thermal reactors, using the resulting plutonium in fast breeder reactors surrounded by 'blankets' of thorium, and finally building reactors fuelled by U-233.

But it now turns out that the programme is stuck at the first stage. By the end of the decade, India will have only ten reactors, fuelled by natural uranium, producing a total of about 2,320 MW, about 1,000 MW more than at present.

K. S. Jayaraman

Budget woes cast chill on French research

Paris. Concern is growing again among researchers at France's publicly financed research institutes about cuts in government support. INSERM, the main agency supporting biomedical research, has just announced that it is freezing 25 per cent of the operating budgets of its laboratories.

It will be known at the beginning of April whether the money will be reinstated. At the same time, the Centre Nationale de la Recherche Scientifique (CNRS) has demanded the return of all the money held by its laboratories at the end of 1995.

The CNRS has been in a critical financial situation for the past two years. It has had to face up to a heavy deficit resulting from the government's failure over several years to provide the funds for programmes it had promised to support.

In 1986, for example, the government of the then prime minister, Jacques Chirac, left unpaid FF181 million (US\$36 million) owed to the research agency. Later, in 1991 and 1992, it was the socialists, faced with economic difficulties after their return to power, who held back a further FF450 million. On top of this were various smaller unpaid sums of about FF50 million. The figures were relatively low compared to the total CNRS budget of about FF13 billion. Nevertheless, the overall deficit of the agency came to about FF1 billion in 1994.

To meet this budgetary crisis, CNRS officials first demanded, in 1994, that its laboratories freeze 40 per cent of their operating budgets. Then, in August 1995, the government agreed to provide CNRS with an additional FF300 million, while at the same time insisting that the agency drop FF200 million of previously approved spending for "very old" programmes. Finally, last autumn, the government promised an extra FF227.7 million in 1996 to cover previous commitments.

At the end of all this, the CNRS's deficit should total only FF270 million this year, of which FF90 million relates to laboratory projects and the rest to the operation of regional offices and the purchase of research equipment and buildings.

In order to eliminate the "very old" programmes, and to revitalize the projects that will benefit from the new money, the directors of the various departments within the CNRS demanded last month the return of all funds remaining in the accounts of individual laboratories. Eventually, the CNRS hopes that it will be able to recoup about FF700 million.

Researchers and their unions are talking about a "hold-up", particularly as the money being reclaimed by the CNRS includes funds raised by individual laboratories through contracts paid for by outside organizations or industrial bodies, as well as from their reserves.

Pierre Tambourin, the director of life sciences at CNRS, and acting director while the permanent director general, Guy Aubert, is in hospital, has promised that the funds raised from outside will be returned. But "those who have made savings will lose them, while those who have overspent — even if in good faith — will have to pay back what they owe". Tambourin adds that the agency "clearly indicated that funds allocated for 1995 had to be spent in full, but that those frozen in 1994 should not be touched".

As a result of these moves, some laboratories now find themselves faced with allocations for 1996 that have been significantly reduced — sometimes to a symbolic one franc. Such is the case, for example, with six out of 28 laboratories in molecular chemistry, and 8 laboratories in the life sciences section. Tambourin admits that this situation should not have occurred, as Aubert had required the repayment of the 'debt' of the laboratories to be made over two years.

But researchers have not taken kindly either to the measures that have been introduced, or to being accused, in particular by Aubert, of being bad managers of their

finances. Jacques Fossey of the Syndicat National des Chercheurs Scientifiques (SNCS) feels, for example, that Aubert has been over-dramatizing the situation in order to build up a "war chest" to enable him to control the future policy of the agency.

Tambourin disputes such claims. But he acknowledges that the director general is keen, independently of budgetary problems, to ensure that the CNRS has a set of large programmes that are more significant than those it has been supporting up to now, costing about FF300 million a year. The funds claimed back last month will be used partly to launch these programmes.

To achieve further cost reductions, CNRS officials are also renegotiating the contracts that link some laboratories to the universities in which they operate. Although most of these laboratories will continue to be linked to the CNRS on the same financial basis as before, between 20 and 30 per cent — those in which the number of university faculty members is higher than the number of CNRS researchers — risk seeing their formal association with the agency weakened, and their funding cut as well.

Catherine Tastemain

Moroccan mineral is truly new and blue

London. **Scientists who were asked during Britain's National Science Week last year to investigate a blue powdery mineral purchased from a roadside stall in Morocco confirmed during this year's Science Week, which began last Friday (15 March), that the mineral (right) has not previously been identified.**

The mineral's owner, Anna Grayson, a geologist and television presenter, enlisted the help of the mineralogy department at London's Natural History Museum after she tried — and failed — to identify it herself. Initial X-ray diffraction tests confirmed that the mineral had not been described before. A team of researchers at the museum then began a series of tests to discover its structure and chemistry.

Viewed under an electron microscope, the mineral contains millions of crystals made up of sub-micrometre-width fibres. Its chemistry was investigated by firing a beam of electrons into the crystals, which energized their atoms to emit X-rays. The mineral's atomic elements — including silicon, aluminium, calcium, magnesium, iron and oxygen — were identified by their characteristic X-ray energies.

The display of the mineral at the museum is one of 5,000 events taking place all over Britain as part of SET

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

'96, the National Week of Science, Engineering and Technology. This is an annual government initiative, coordinated by the British Association for the Advancement of Science and aimed at increasing public awareness of science.

The week is supported by industry, universities, schools and the media. Government ministers are using the occasion to promote both science itself and their current policies on how it should be supported. Opposition political parties have also lent their support, while pointing out that the UK government's spending on research and development has declined for the past decade.

Ehsan Masood

Panel calls for overhaul of AIDS research

Washington. A lengthy, long-awaited and often critical report by a 118-member panel of scientists and other experts has called for a major revamping of the \$1.4-billion AIDS research programme of the National Institutes of Health (NIH), which it says has been poorly focused and uncoordinated.

The report recommends that NIH scientists be stripped of strategic control over vaccine research, and that funding for AIDS research by non-NIH scientists be doubled. The new money would come in part from work wrongly labelled as AIDS-related. The report also wants drug discovery efforts scaled back, on the grounds that many duplicate the drug industry's work.

At the same time, the report calls for more basic research, especially in immunology, prevention and opportunistic infections, in a programme that consumes 12 per cent of NIH's budget. It suggests that a mish-mash of clinical trials be consolidated into a single network, and describes as "critical" the need actively to recruit promising and distinguished scientists to a field that consumes 12 per cent of NIH spending.

The report, whose executive summary was released last week, also says that it is "crucial" for the NIH's Office of AIDS Research (OAR) to keep control over AIDS research at the NIH's 24 institutes and centres — a role that has recently been challenged by congressional Republicans. And it recommends against establishing a separate AIDS institute at the NIH.

The report is "an excellent mid-course correction" to NIH's 15-year-old AIDS research effort, says Arnold Levine, professor of molecular biology at Princeton University, who chaired the AIDS Research Programme Evaluation Working Group. The group included 90 scientists, as well as activists and representatives of the biotechnology and pharmaceutical industries.

The conclusions of the working group were unanimously accepted by the advisory council of the OAR, which commissioned the report in late 1994. Harold Varmus, the director of NIH, said that he and individual institute directors will begin consulting on responses immediately — a reaction the report calls "imperative" in order to influence budgets for 1997 and 1998.

"We will be taking these recommenda-

tions with extreme seriousness," Varmus said last week. Implementation "will take some time, but will be effective".

The report marks a watershed in AIDS research at NIH, which accounts for 85 per cent of public spending on AIDS research worldwide. While commending past NIH efforts for producing "unprecedented dividends" that have created "the first real chance" of turning AIDS from an inexorably fatal condition into a chronic, manageable disease, it is also highly critical of many aspects of that research.

The overarching theme is the need for better coordination, prioritization and focusing of AIDS research, in which all 24 NIH institutes and centres are involved, of which the National Institute of Allergy and Infectious Disease (NIAID) consumes 42 per cent of AIDS money, and the National Cancer Institute 16 per cent.

One key recommendation is that the OAR and the institutes develop stricter definitions of AIDS and AIDS-related research. At many institutes — in particular the NCI — a "substantial" proportion of AIDS money continues to be funnelled to activities "with little or no direct relevance" to AIDS. As an example, it cites programmes to develop artificial blood substitutes by the National Heart, Lung, and Blood Institute — the US blood supply has been protected from HIV since 1985.

The NIH also comes under fire for holding the AIDS grant purse strings too closely, channelling "simply insufficient" funds to unsolicited proposals from external scientists, in contrast to projects proposed and controlled by NIH scientists. In 1994, for example, about 20 per cent of NIH spending on AIDS research went to unsolicited proposals, compared to 50 per cent of money for non-AIDS research.

As a result, the report says, outside investigators are discouraged from proposing novel AIDS-related experiments. As a partial remedy, it calls for funds for investi-



Levine: sees report as 'excellent correction'.

gator-initiated research to be doubled, the new money coming partly from research now funded by AIDS dollars but "only peripherally related to AIDS".

Another significant recommendation is that the entire AIDS vaccine research effort be restructured, given that vaccine research has "received less funding and attention" than other areas at NIH, and that the spread of HIV worldwide means this is no longer acceptable. "In many developing nations, vaccines may be the only cost-effective way to prevent transmission," the panel writes.

In particular, it recommends establishing a trans-NIH vaccine research effort as an independent unit within the NIAID. The effort would be directed by an AIDS Vaccine Research Committee (AVRC), chaired by and composed primarily of non-government scientists. Vaccine research should return to fundamentals in the light of recent failures, Levine adds. "There haven't been enough attempts to explain why they failed," he says. "That's basic science."

The panel suggests that all AIDS clinical trials be condensed into a single network sponsored primarily by NIAID and overseen by an OAR committee to ensure coordination between the institutes. It is particularly critical of the NCI's Developmental Therapeutics Program (DTP), calling it of "limited" productivity in producing novel agents and largely replicating work by the pharmaceutical industry. Restructuring and "substantial" cuts in its AIDS funding for the DTP are "appropriate".

AIDS activists have been divided in their reaction to the report. Some have given it a warm welcome. "The nation's leading scientists have provided NIH with a clear roadmap," said Mark Harrington, a member of the working group's executive panel, and a policy analyst with a New York AIDS group.

But other activists criticized the report for not going far enough. "It completely fails to address several of the much more fundamental flaws in NIH decision-making and priorities," says Bob Lederer, a spokesman for ACT UP/New York. He complains that the report does not recommend targeting specific money to alternative therapies, that it does not recommend conflict of interest rules for NIH AIDS scientists with links to drug companies, and that it fails to call for research targeted on the needs of women and non-whites.

In addition to a 17-member executive working group, the full panel was made up of six subject review panels, which addressed the areas of aetiology and pathogenesis, drug discovery, clinical trials, vaccine research and development, behavioural, social sciences and prevention research, and natural history, epidemiology and prevention research. Their reports will be released early next month.

Meredith Wadman

Clinical trials launched of Salk AIDS vaccine

Washington. Immune Response Corporation, the California biotechnology company co-founded by Jonas Salk, the late polio vaccine pioneer, announced last week that it has launched a phase-3 trial of the controversial therapeutic AIDS vaccine Remune, based on an inactivated version of the AIDS virus.

The company says that it will enlist at least 2,500 patients at more than 50 US medical centres in an attempt to establish whether Remune delays the onset of AIDS in HIV-positive individuals. The Food and Drug Administration gave final approval for the phase-3 trials to proceed earlier this month.

M. W.

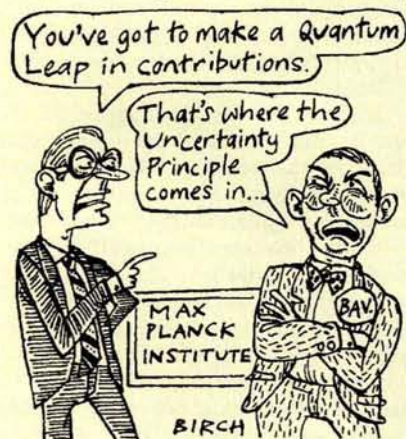
Bavarian discontent threatens funding deal for institutes

Munich. Bavaria, home to some of Germany's largest research centres, is threatening to pull out of a long-standing national agreement on the financing of research institutes if the federal government backs a plan requiring each *Land* (state) to pay a higher share of the costs of the Max Planck institutes it hosts.

This could mean renegotiating the funding systems for all of Germany's research organizations, including the Fraunhofer Society and the Max Planck Society (MPG). At present, the costs of each organization are divided between Germany's federal and 16 *Länder* governments, according to a fixed ratio.

In the case of the MPG, this ratio is 50:50. But, unlike the other three organizations involved, the MPG does not allow a host *Land* to pay the full *Land* share of an institutes' costs, so as to maintain its political independence. Instead, each *Land* pays into a common fund used to finance half the *Land* share of all MPG institutes. So a host *Land* pays directly only a quarter of the costs of its own institutes, limiting its influence in such an institute's fate.

Länder with few such institutes resent subsidizing those in other *Länder*,



and earlier this month *Länder* prime ministers agreed at a meeting in Berlin to abandon the common fund by 1999. Three *Länder*, Bavaria, Baden-Württemberg and Berlin, each with relatively large numbers of institutes, would face hefty bills, and voted against the move.

The decision has still to be approved by the federal government. But Bavarian prime minister Edmund Stoiber told the Berlin meeting that, unless a compromise is found, Bavaria may pull out of all joint funding agreements from 1999, forcing a major revision of research funding.

A.A.

Research centres face chaos after German ruling on tax

Munich. State-funded research institutes in Germany have been thrown into turmoil by a tax ruling that could require them to register as private companies, and severely limit their access to public funds.

The ruling was made by a federal taxation court in Munich, which decided that the DLR (Deutsche Forschungsanstalt für Luft- und Raumfahrt), the national research centre for air and space research, must pay value added tax (VAT) on industrial research contracts at the full 15 per cent rate applicable to commercial companies. Organizations, including publicly funded research institutes, that do not make a 'profit' on the research they do, pay only 7 per cent.

It is more than 20 years since tax inspectors first demanded that DLR pay the tax at the full rate. The research centre refused to do so, while putting aside the difference, amounting to DM36 million (US\$25 million), in the hope that it would be allowed to keep these funds. Since 1985, the centre has been paying the full 15 per cent, hoping to win back excess payments in the courts.

But a lengthy legal battle, which started with the local finance office in Cologne in 1990, has now been lost. Last week DLR's case was rejected by Germany's highest financial court, requiring the centre to pay back the DM36 million. In theory, the decision could require all research institutes that carry out work for industry to raise the price of their contract research. This would seriously undermine their competitiveness; indeed, Erich Maaß, a member of the education and research working group of the Christian Democrats, the senior coalition partner in the Bundestag, points out that it would conflict with the government's wish to see closer links between academic institutions and industry (see *Nature* 4, 372; 1994).

There could be other consequences, too. First, the ruling might allow the government to demand back payments for tax from other applied research institutes. The KFA in Jülich, for example, which specializes in energy and environmental research, could face a bill of around DM20 million.

Second, in order that the tax may be collected, all research institutes with any income from industry would have to be legally defined as profit-making companies. This would make them liable for full VAT on all income — public as well as private, on basic as well as applied research — as well as various other taxes from which they are at present exempt. According to a spokeswoman from the Max Planck Society, for example, this would cost the society thousands of millions of marks per year.

The most dramatic consequence, according to Klaus Fleischmann, managing direc-

tor of the Helmholtz Society, representing Germany's 16 national research centres, is that as private companies, the institutes would, under European Union competition laws, be allowed to receive no more than 25 per cent of their income from government.

This would mean that the financing of all public research institutes — including national research centres, Max Planck institutes, the applied research institutes of the Fraunhofer Society, and the 'Blue List' group — would need to be restructured, as they rely primarily on support from national and local governments. "The traditional system is threatened," warns Maaß.

No-one wants this to happen. Not only would it mean less money for research, but, as Hans-Ulrich Wiese, a member of the Fraunhofer Society board, points out, it would undermine the government's efforts to boost technology transfer. For example, if research institutes lost their status as public bodies, they would be unable to use staff paid out of public funds to carry out contract research, which would force up costs even further than the increased VAT.

Research organizations also claim that it is incongruous for the government to take back with one hand (the finance ministry) what it has given out with another (the research ministry). Beatrix Vierkorn-Rudolf, who runs the WBL, the newly created umbrella organization for Blue List institutes, finds it "absurd" that "at a time when the catch phrase is 'lean government', money should be shifted pointlessly between different ministries".

Although the court's decision was confirmed only last week, research organizations have been meeting with officials from the research ministry for several months to find a speedy way out of their dilemma.

A working group made up of representatives of these organizations, and headed by Fleischmann, claims that the only possible solution is a change in the law that would allow all research institutes to be considered as public bodies, irrespective of whether they have income from industry.

Such a move is supported by the Christian Democrats' education and research group. And last Monday (18 March), a group of expert officials from the finance ministry announced that they would be recommending to the finance minister that he should also support the idea. But, even if it were to be adopted by the government, it could not come into effect until the beginning of next year. Up to then, says Fleischmann, the finance ministry should institute a '*Nichtanwendungsbeschluss*' — meaning that the court ruling should not be enforced.

Alison Abbott & Quirin Schiermeier

Science loses cabinet seat after Australian election

Sydney. Australia's new conservative Coalition government has announced changes in its science policy arrangements following the general election at the beginning of this month. The Liberal and National Parties, linked in a coalition led by John Howard, won a resounding 46-seat majority over the 13-year-old Labor administration of the former prime minister, Paul Keating.

The new science and technology minister is Peter McGauran, a member of the National Party, which captured the rural vote. Unlike his predecessor, Peter Cook, who had a seat in the cabinet as a full Minister for Science, McGauran will be a junior minister in the large Department of Industry, Science and Tourism (DIST), and will attend cabinet meetings only when invited.

Some scientists are concerned that their activities are no longer directly represented in the cabinet. But few will be totally surprised at the new government's general approach, as McGauran is following a 20-page policy for science, engineering and technology, *Investing in Tomorrow*, which was issued in the final two weeks of the election campaign by Robert Hill, the coalition's spokesman on education and science.

The period since the election has heard a predictable outcry from the new government that the economy is in worse shape than it had realized, and that \$A8 billion (US\$6 billion) of expenditure has to be cut. Nevertheless, McGauran said last week that the government "will honour and implement all our commitments [in science], notwithstanding the deficit we inherited".

The coalition's financial promises total \$A190 million, including restoring \$A20 million per year to the triennial baseline funding of the Commonwealth Scientific and Industrial Research Organization, the country's main research agency, and an extra \$A90 million over three years for research infrastructure in universities. Funds for the

Australian Research Council's Collaborative Research Grants will increase by \$A30 million and postgraduate scholarships by \$A9 million over the same period.

Within the government itself, two bodies are being transferred to an enlarged Science Branch within DIST: the Prime Minister's Science and Engineering Council (PMSEC), an advisory body set up by former Labor Prime Minister, Bob Hawke, and the Australian Science and Technology Council, which will have engineering added to its name (to become ASTEC) and increase its role in analysing scientific performance.

Another significant shift, not hinted at in pre-election statements, has been the government's decision to move the post of chief scientist from the Prime Minister and Cabinet Department to DIST, and at the same time to enhance "the importance and status of the chief scientist with legislative authority to fulfil a cross-government role".

McGauran says he is consulting widely about the appointment of a chief scientist to succeed Michael Pitman, whose final term of office has come to an end. He points out that both PMSEC and the chief scientist will still have direct access to the prime minister, although observers see the moves as equivalent to a political downgrading. ASTEC, however, will have its own secretariat and will become "independent of government".

The Australian Research Council, which is at present part of the Education ministry and is responsible for providing competitive grants primarily to university researchers, will become a "separate body, quite independent of government", says McGauran. But there are signs that the council may become McGauran's responsibility, giving him what could be the most centralized role in the administration of science in Australian history, even though he will be lower in the ministerial pecking order than his predecessor.

Peter Pockley

Resignation defuses Witwatersrand row over immunologist

Cape Town. A fierce dispute between a black university administrator and faculty members at the University of the Witwatersrand in South Africa appears to have been defused by the decision of William Makgoba, a prominent research immunologist, to resign as the university's deputy vice-chancellor.

The move follows the signing of a 'memorandum of agreement' between Makgoba and nine of thirteen academics — including ten deans — who last year accused him of embellishing his curriculum vitae, carrying out his administrative duties inadequately, and bringing the name of the university into disrepute (see *Nature* 378, 324; 1995).

The memorandum, which was accepted by the university's council last week, recommends that a tribunal that was to have investigated the claims against Makgoba should be scrapped. He will occupy a research professorship in the health sciences faculty for the remaining three and a half years of his contract.

The agreement between Makgoba and his critics follows the conclusion of a commission headed by Malcolm Wallis, chairman of the Senior Council of the Bar of South Africa, that Makgoba's counter-allegations of tax evasion by the thirteen were groundless. In the memorandum, Makgoba apologizes for publicizing and misrepresenting information from their personal files.

In turn, the nine signatories to the memorandum acknowledge both that Makgoba "has made a significant and internationally-recognised contribution to immunology", and that the university had given him "inadequate support" in overcoming his relative lack of administrative and managerial experience. Makgoba, a South African who left the country after graduating in 1976, was recruited by the university 18 months ago from the Royal Postgraduate Medical School in London, where he was a senior lecturer.

The four remaining academics submitted a separate memorandum to the council, and have released a joint statement. They are unhappy about both the decision to scrap the tribunal and the fact that the memorandum does not exclude Makgoba from becoming the university's next vice-chancellor when the present incumbent, Robert Charlton, retires next year.

But this outcome now appears unlikely. In justifying his resignation, Makgoba stated that he could not have worked with ten deans who did not want to work with him. He added that he knew "that I had to leave this office because I would not have been effective any longer". Michael Cherry

NATURE · VOL 380 · 21 MARCH 1996

Telescope prospects looking dimmer

Sydney. Peter McGauran, Australia's new science minister, warned last week that astronomers should not take for granted that the Coalition government will fully back their efforts to join the European Southern Observatory (ESO) project to construct and operate the Very Large Telescope (VLT) in Chile.

The hopes of astronomers had been raised during the election campaign that their application might be supported, whichever party won. Despite its rejection in the Innovation Statement published last December, Peter Cook, the former science minister, said that fund-

ing the ESO application to the extent of A\$30 million (US\$23 million) over five years would have top priority for Labor in the August 1996 budget (see *Nature* 379, 668; 1996).

At the time, the opposition spokesman, Robert Hill, said he was happy for negotiations with ESO to continue. But Hill stopped short of matching the then-government's financial commitment. McGauran now says he "would seek independent advice of the size of the investment" before making any decision. "I warn against unreal expectations in the present economic climate." **P. P.**

Tectonic rift 'may threaten' Lesotho dam

Cape Town. Controversy is raging in South Africa about a series of earthquakes that have taken place at a new dam in Lesotho since it started to be filled last October. A prominent geologist is claiming that the dam may be more dangerous than the government accepts, as it may lie on an incipient boundary between two tectonic plates.

The Katse dam forms part of the Lesotho Highlands Water Project, a joint venture between Lesotho and South Africa, planned to provide both hydro-electricity and water for the two countries.

The largest of nine seismic events, with a magnitude (M) of 2.5 on the Richter scale, was recorded on 3 January. A crack subsequently appeared stretching from the edge of the reservoir through the village of Ha Mapeleng to a 2,350-metre peak above it. The fracture is now 1.5 km long, and wide enough to insert a broom-handle.

With a height of 185 metres, the Katse dam will ultimately hold back about two billion tonnes of water, a load which, according to one model, will depress the surface of the Earth by about 3 cm over an oval-shaped area about 6 km upstream of the dam wall, near Ha Mapeleng.

The possibility of dam-induced seismic activity was pointed out by the original impact assessment for the project, carried out in 1988. The Geological Survey, now the Council for Geosciences, was commissioned to maintain three seismographic stations in Lesotho to monitor the situation. But the recent seismic activity has triggered a sharp debate between Chris Hartnady, professor of geology at the University of Cape Town, and the South African Department of Water Affairs and Forestry, which is responsible for the South African side of the project.

The disagreement focuses on the maximum size of an earthquake likely to occur at Katse. The dam has been built to withstand an $m6.5$ earthquake. But Hartnady, suggesting that the entire area may lie in an intermediate zone between two plates of the Earth's lithosphere, claims that it should be required to withstand up to $M7.2$.

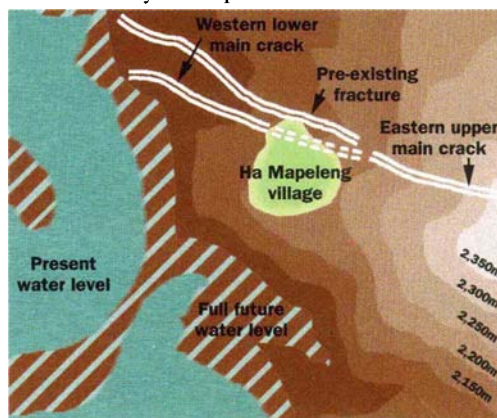
The East African Rift valley — a zone running from the Jordan River through Africa to the Zambezi — where the African continent is being extended, is widely thought of as representing the early stages of the break-up of Africa into two smaller plates. In 1990, two years after the initial impact assessment for the dam was carried out, Hartnady published an article in the *South African Journal of Geology* speculating that this incipient plate boundary might extend south into the Natal/Kwazulu-Lesotho-Free State area.

If such speculation proves correct, the dam site would lie not at a seismically stable position inside a single plate, but within a wide boundary zone between the Nubian

and Somali plates (see figure, below).

The seismicity associated with this postulated East African Rift propagation is well known: the two largest African earthquakes this century were the Rukwa event in Tanzania in 1910, and the Juba event of 1990 in the Sudan with magnitudes of 7.3 and 7.2 respectively.

The official calculation that the maximum likely earthquake in the Katse dam



The Katse dam lies on what may be an extension of an incipient boundary between two tectonic plates.

region would be $M6.5$ is based on plotting the frequency of earthquake occurrence and magnitude over time. But Hartnady claims that this is "tantamount to assuming that the largest earthquake that has occurred within the relatively short interval of instrumental recording is roughly equivalent in size to the largest earthquake that can be expected in the future".

Hartnady says there is an urgent need for a tectonic assessment of the crustal stress. A realistic seismotectonic model, incorporating Nubian-Somalian plate-kinematic constraints and earthquake experience from similar zones elsewhere in the East African Rift System suggests, he claims, that the size of the maximum credible earthquake in the Katse area is not less than $M7.2$.

"Contrary to the previous perception of material homogeneity, the Earth's crust between and around the Katse reservoir contains a wide variety of structural discontinuities which are susceptible to neotectonic reactivation in the modern crustal stress field," says Hartnady. "This could provide a focus for reservoir-triggered earthquakes."

Willie Croucamp, the managing engineer responsible for international projects at the department, disagrees. He points out that Hartnady is invoking the idea of a wide boundary zone because of uncertainty as to whether the Rift valley extends into South Africa or not.

But Croucamp says the uncertainty about Hartnady's own theory means that it cannot be used for seismic risk assessment, emphasizing that the value of $M6.5$ was calculated according to guidelines laid down by the

International Commission on Large Dams.

Two seismic disasters in the vicinity of reservoirs have occurred in the past 30 years, both in India. In 1967, a reservoir-induced earthquake measuring $M6.5$ at the Koyna Dam, located in a previously aseismic region, killed 177 people and injured 2,300 others. In 1993, an $M6.2$ earthquake, concentrated between 10 and 20 km from a reservoir near the town of Khillari, resulted in almost 10,000 deaths and 30,000 injuries — although it has not been proved that the earthquake was triggered by the construction of the reservoir.

A report on the recent seismic events commissioned from two outside experts by the Lesotho Highlands Development Authority, and released at the end of February, says that there is no evidence associated either with the observed cracks or with the surrounding geological features to suggest that the village or any large rock mass near the reservoir is in danger of sliding into it. (The latter is what happened at Vaiont in Italy in 1963, when a 300-million-tonne avalanche fell into the reservoir, producing a towering wave that killed nearly 2,000 people.)

The report predicts that shallow earthquakes, even of the order of $M4.5$, could continue as the dam is filled, and recommends rehousing the villagers in earthquake-resistant houses — some of the villagers' huts have already been damaged. But the report supports the original 1988 assessment that the maximum likely earthquake at Katse is $M6.5$. "There is no evidence from the activity that has occurred [to] suggest that a deep-seated earthquake is imminent," the report concludes.

Nevertheless, the department has invited Hartnady to set out his views in a formal submission, and has agreed that his evidence and claims should be reviewed by a panel of experts, a process likely to take two months.

The local villagers, meanwhile, have an alternative explanation for the earthquakes at Katse. They believe the tremors are a manifestation of the wrath of the river serpent, which has been confined by the dam and is now unable to move up and down the length of the river.

Michael Cherry

Einstein fails to sell

New York. Albert Einstein's original 72-page manuscript that set out to unify mass and energy failed to sell at an auction in New York last week. The highest bid — \$3.3 million — is said to have fallen short of the manuscript's minimum asking price. The auction house selling the manuscript, Sotheby's, had previously suggested that the document, which illustrates how Einstein developed some of his key ideas, could fetch up to \$6 million. □

Russia pays researchers' salaries at last, but science still ailing

Moscow. The Russian government has belatedly paid up its 266 billion rubles (US\$55 million) owing to the Russian Academy of Sciences (RAS) to cover unpaid salaries for academy scientists for the past three months. But the scientists fear that the payment, as well as an additional R5 billion 'goodwill bonus', will still be insufficient to reverse the decline in Russia's scientific fortunes.

Yuri Osipov, president of the academy, has described the money as a "social allowance", and warns that Russian science "will die" unless its financing is "substantially increased". Addressing a press conference in Moscow last week, he said that the material resources of the academy's research institutes were "totally exhausted". Osipov added: "The money we have got should be considered as social allowance rather than salary, as it is practically impossible to carry out scientific research."

Valentin Koptug, chairman of the Siberian branch of the academy, said that despite a sevenfold drop in funding over the past four years, he had lost only 8,639 out of 66,102 research staff. The academy was devoting all its energy to ensuring that it did not "miss the major target of the day: to preserve the scientific environment in the country until better times arrive", said Koptug. □

UK embryo authority apologizes

London. Britain's Human Fertilization and Embryology Authority (HFEA) has apologized to a group of French researchers for wrongly claiming that they had withdrawn a study suggesting that mice born from frozen embryos had differences in morphology and behaviour.

The HFEA published a report in December last year which suggested extending to ten years the period allowed for the storage

of embryos. The report cited an unnamed "French study" whose conclusions, it alleged, "have been withdrawn" by the authors. But a spokeswoman for HFEA said that the authority now acknowledges that the conclusions of the study by Pierre Roubertoux, professor of genetics, neurogenetics and behaviour at the University of Paris, and his colleagues were not withdrawn.

After the publication of his results, Roubertoux acknowledged that none of them gave cause for concern about the safety of embryo freezing techniques (see *Nature* 373, 553; 1995). The researchers found that cryopreservation caused no "major anomalies". □

Japan links up with Mir in space

Tokyo. The National Space Agency of Japan (NASDA) and the Russian Space Agency (RSA) signed an agreement on 12 March, under which RSA will facilitate the implementation of NASDA experiments on board the Russian Mir space station. NASDA is carrying out these experiments as part of its preparations for the completion of its own Japan Experimental Module (JEM) for the US space station.

Under the agreement, which NASDA says is the first between the two agencies, RSA will facilitate experiments on long-term monitoring of cosmic radiation and the cellular and genetic effects of such radiation. Studies will also be undertaken on the growth of microflora in microgravity environments, in order to obtain data relevant to the prevention of microorganism contamination of on-board equipment and maintaining a healthy environment. □

Boost for Russian Internet access

Moscow. George Soros, the US philanthropist, has agreed with the Russian government to provide access to the Internet for 30 Russian universities located outside Moscow and St Petersburg.

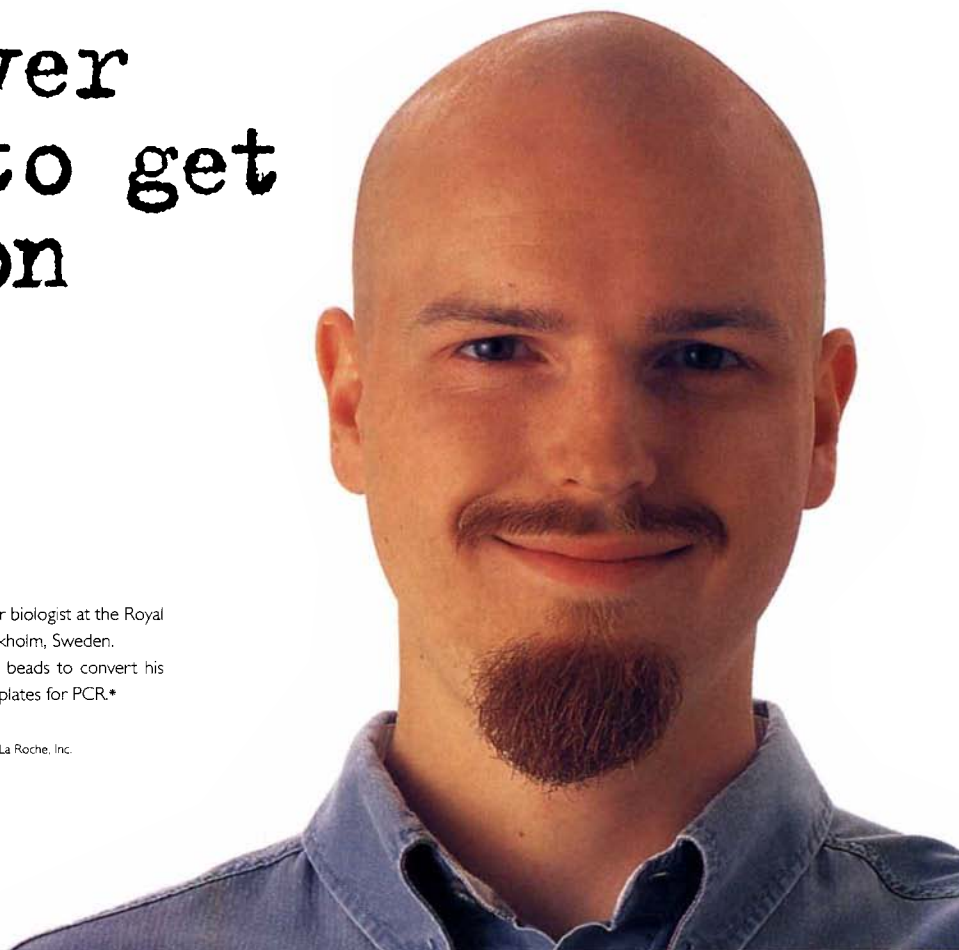
Following a meeting with Viktor Chernomyrdin, the Russian

Patrik never fails to get a reaction

Patrik Samuelson is a molecular biologist at the Royal Institute of Technology in Stockholm, Sweden.

Patrik uses Ready-To-Go beads to convert his RNA samples into cDNA templates for PCR.*

* PCR is a patented process of Hoffmann-La Roche, Inc.



What did Heisenberg know?

SIR — Regarding the question of how much Werner Heisenberg knew about the explosion of the atomic bomb (book review by Irving M. Klotz, *Nature* 379, 411; 1996), I have a — possibly exaggerated — statement by him in his own handwriting. In Germany and Austria, there was a rumour after the dropping of the bombs on Japan that Nazi Germany had invented the atomic bomb but had not used it and that the Americans had appropriated this German “miracle weapon”.

Although this was clearly untrue, I took the opportunity of a visit by Heisenberg to Vienna in the early 1950s to interview him and to write an article entitled “Professor Heisenberg: Hitler’s Germany had no atomic bomb. An authoritative statement by the head of German atomic research”. (I was at the time science editor of the Viennese newspaper *Arbeiter-Zeitung*.)

Heisenberg checked my copy but changed hardly anything, because, on his advice, I had based my article on a paper by him in *Die Naturwissenschaften* (33, 325; 1946). But at one place he added a sentence that reads: “*Man wusste aber in Deutschland, wie man die Atomenergie im ‘Reaktor’ zum Betrieb von Maschinen verwenden kann, man wusste auch, dass im Reaktor der Sprengstoff Plutonium (der in Deutschland nicht so hiess) gewonnen werden kann, aus dem Atombomben gemacht werden.*” (“But it was known in Germany how [the] atomic energy from the ‘reactor’ can be used for the running of machines; it was known too that the explosive plutonium (which in Germany was not known by that name) of which atomic bombs are made can be generated in the reactor.”) Heisenberg wanted to add something further but crossed out the first word.

Contrary to Heisenberg’s statement and as Irving M. Klotz says in his book review, it is doubtful whether much was known about plutonium in Nazi Germany. In 1944 (with a preface dated September 1943), the book *Einführung in die Kernphysik* by Wolfgang Riezler was published in Leipzig. As the text on transuranium elements on page 161 shows, a small quantity of element 93 (neptunium) had been produced, but element 94 (plutonium) had not yet been isolated:

Irradiating uranium with a strong neutron source, it was possible to produce enough of the element 93 from U^{239} ... The nucleus 93^{239} also is a beta-emitter with a half-life of 2.3 days. It must transform, therefore, by its radiation to an isotope of element 94. It was not possible, however, to discover radiation related to this new nucleus although uranium was exposed for a whole year to very strong neutron radiation produced by means of

a cyclotron. It is therefore supposed that this new nucleus is a very long-living radioactive substance, probably an alpha-emitter. Its life-span has to be more than a million years [in fact, it is 24,390 years, E.K.], otherwise its alpha-radiation would have been detected by now.

Of course, secret research could have been further advanced than Riezler knew, but Allied bombing and work of (in German eyes) more military importance did not leave much opportunity for it.

What Klotz says is corroborated by Heisenberg in his paper in *Die Naturwissenschaften* about the German nuclear energy efforts during the war (p. 327): “Furthermore, it was to be expected that one can produce an explosive for atomic bombs in a uranium burner. [The term ‘reactor’ was not yet used, E.K.] However, no investigations about the technical side of the atomic bomb problem, e.g., about the minimum size of a bomb, had been made.”

Friedrich Katscher
Mariahilfer Str. 133,
A-1150 Vienna,
Austria

Pointed error

SIR — I have often enjoyed Birch’s satirical cartoons on topical scientific issues. He frequently includes clever symbols in his cartoons to emphasize his point. In the past few years, he has included the Saguaro cactus in drawings depicting events that have occurred in the Mojave Desert of California and in central New Mexico (*Nature* 364, 750; 1993 & 377, 96; 1995).

Finding a Saguaro cactus in either of these locations would be a considerable range extension for this plant. The Saguaro cactus is endemic only to the North American Sonoran Desert. This limits its distribution in the United States to southern Arizona and extreme south-eastern California. Saguaro cacti would not naturally be found in either of the locations where Birch has placed them.

Wayne A. Van Voorhies
Department of Molecular
and Cellular Biology,
University of Arizona,
Tucson,
Arizona 85721, USA

Correspondence

Letters submitted for Correspondence should be typed, double-spaced, on one side of the paper only, or e-mailed to nature@nature.com

Dirac was not an “obdurate” atheist

SIR — In a report of the unveiling of a plaque honouring P. A. M. Dirac in Westminster Abbey (*Nature* 378, 223; 1995), it is stated that he was an “obdurate atheist”.

This may be true, but in an article entitled “The evolution of the physicist’s picture of nature” (*Scientific American* 208, 45–53; 1963), Dirac said:

It seems to be one of the fundamental features of nature that fundamental physical laws are described in terms of a mathematical theory of great beauty and power, needing quite a high standard of mathematics for one to understand it. You may wonder: Why is nature constructed along these lines? One can only answer that our present knowledge seems to show that nature is so constructed. We simply have to accept it. One could perhaps describe the situation by saying that God is a mathematician of a very high order, and He used very advanced mathematics in constructing the universe.”

This does not sound as if it was written by an atheist, at least not an “obdurate” one.

Fernando Orrego
Faculty of Medicine,
Universidad de los Andes,
Casilla 20106,
Santiago 20,
Chile

Bieterminology

SIR — ‘Bieterminology’ exercised the minds of R. P. J. Swannell *et al.* and Calvin Dytham (*Nature* 378, 14; 1996). I offer solutions to both their problems.

The idea of reserving the term ‘bioremediation’ for those acts that aim to enhance the intrinsic biodegradation of contaminants (Swannell *et al.*) has my support. Licensing agencies would presumably like to pay particular attention to processes that depend on the addition of extrinsic biota, genetically manipulated or otherwise, and for this the term ‘bioaugmentation’ is appropriate.

As for the slowed increase in the rate of papers with “biodiversity” in the title (Dytham), those of us who turned our attention away from taxonomy in the past because of the indifference of funding agencies are embittered. I suppose we should now stop instructing our students to exterminate any specimens not described in the field course guide.

Alan J. McCarthy
Department of Genetics
and Biology,
University of Liverpool,
Liverpool L69 3BX, UK

A big step along the visual pathway

David Hubel

The results of some technically demanding experiments may resolve a long-standing debate — the cause underlying the orientation selectivity of cells in the primary visual cortex.

In their paper on page 249 of this issue¹, Ferster, Chung and Wheat come close to establishing the circuit that forms the basis of the simple cells of the cat primary visual cortex. The problem they have tackled — the mechanism whereby these cells show a phenomenon known as orientation selectivity — has been the subject of dispute for well over 30 years, but to my mind these new results provide the clearest-cut answer yet.

The visual system of mammals is organized as a series of neurally interconnected stages, beginning in the retina and continuing, in the brain, in the lateral geniculate body and then a succession of stages in the cerebral cortex (see figure). The first of these cortical stages is the primary visual cortex. The visual input to this cortical area is relayed from the retinas through the lateral geniculate bodies, and lateral geniculate cells respond best when the retina is illuminated by circular spots of light. Cortical simple cells, by contrast, respond to long narrow lines whose optimal orientations vary from cell to cell. The new work¹ is the latest in a series of studies pointing to the conclusion that this orientation selectivity is the result of direct excitatory convergence of inputs from the lateral geniculate cells, rather than of inhibition within the cortex itself.

Fibres of the optic nerves, and the retinal ganglion cells from which they arise, are already two steps removed from the retinal receptors, the rods and cones. Each ganglion cell receives connections, through one or two intermediate steps, from a close-spaced and roughly circular group of hundreds or thousands of receptors that make up the receptive field of that cell. In 1953 Kuffler observed² that in cats the ganglion-cell receptive field is not uniform, but consists of a small circular central region surrounded by an annulus. In about half the cells, the centre is excitatory: when illuminated it produced brisk machine-gun-like discharges from the cell; illuminating the surround caused a suppression of spontaneous firing. A large spot covering both centre and surround produced only a weak discharge or none

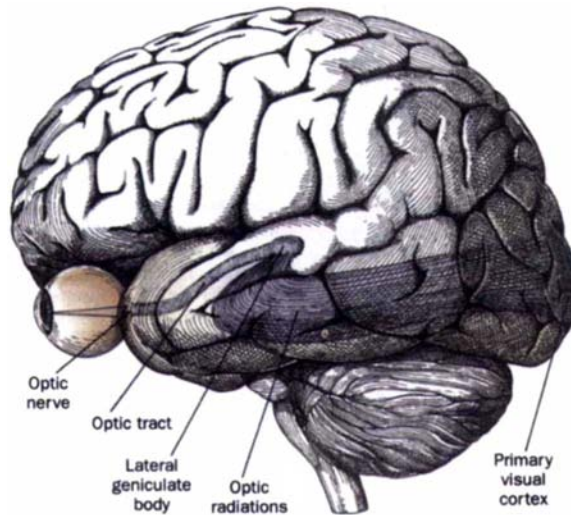
at all. The other half of the ganglion cells ('off-centre cells') worked in just the opposite way, with an inhibitory centre and excitatory surround. Each ganglion cell was thus concerned not with general retinal illumination but with a comparison of the intensity of the light falling in one small region of retina with the intensity in the immediate surround.

This set of discoveries at once explained a number of perceptual phenomena, not

an edge boundary. A cell fires to a line in some specific orientation, vertical, horizontal or oblique, but not at all to orientations more than about 20–30° from the optimum. Cells in the lateral geniculate body, in contrast, had receptive fields that were at least qualitatively identical to Kuffler's ganglion cell fields, with no hint of orientation selectivity.

We found these results exciting because they were the first direct indication that cortical cells perform transformations on the incoming information — in short that the cortex actually does something interesting. Clearly, cells of the primary visual cortex were analysing a visual scene by asking whether any small region contained a light-dark contour and, if so, what its orientation was. That information was then presumably handed on to subsequent stages for further processing, locally and in other visual cortical areas.

It then gradually became clear that these orientation-selective cells were of several types⁴, differing mainly in their complexity. The simplest, which we (imaginatively) called 'simple', had receptive fields that, like retinal ganglion and geniculate fields, were subdivided into excitatory and inhibitory regions, but the geometry was strikingly different. Most often the field consisted of a long, narrow excitatory region flanked on either side by an inhibitory region, so that a bright line confined to the excitatory region produced a brisk response. The width of such an



Visual pathway in the human brain, running from the retina through the optic nerve and tract to the lateral geniculate body (or nucleus) in the thalamus. The visual signal is then passed to the primary visual cortex at the back of the brain by optic radiations. It is there that cells are orientation selective — that is, individual cells respond to a line in some specific orientation but not to lines some 20–30° from the optimum — although the system is slightly different in humans and cats. (Adapted from ref. 11; artwork by Carol Donner.)

least the fact that 'black' has little to do with the amount of light coming from an object, and everything to do with a comparison of that light with light from the immediate surroundings. It explains why a TV set when turned off looks white or grey but not black, whereas we see excellent deep blacks in a normal TV picture.

In the late 1950s, Torsten Wiesel and I began to use these techniques of visual stimulation and single-cell recording in the primary visual cortex of cats³. We soon discovered that most or all the cells in this area respond grudgingly or not at all to small spots of light, but respond with brisk repetitive firing to short lines — a bright line on a dark background, a dark line, or

excitatory region was generally about the same as the diameter of a geniculate receptive-field centre, so we proposed that the simple cell might be receiving direct excitatory input from a group of on-centre geniculate cells whose field centres fell along the excitatory region. The inhibitory flanks could be similarly the result of input from geniculate off-centre cells. Other circuits were clearly logically possible, but this seemed by far the simplest and it found support in the observation that simple cells are grouped together in the cortical layer (layer 4) that receives the geniculate inputs.

Soon, however, several alternative ideas sprang up concerning the circuit

underlying simple cells in cat cortex. Inhibitory synapses, employing the transmitter γ -amino butyric acid (GABA), are plentiful in the cortex, especially in layer 4, and much effort has been expended in trying to understand their function. One of the most popular notions, advanced by Blakemore and Tobin⁵, was that a cell's orientation selectivity might be sharpened if it were to receive inhibitory inputs from other cortical cells with different orientations. This process would be analogous to events in the geniculate, where the power of the receptive field's surround to countermand the influence of the centre is enhanced, and seemed to be confirmed by Sillito⁶ — he observed a loss of orientation selectivity of simple cells when the drug bicuculline, a GABA antagonist, was applied to the cortex.

On the other hand, evidence for the original idea of direct excitation came from Chapman, Zahs and Stryker⁷. They first determined the orientation selectivity of cells in a single orientation column in cat cortex. Then, by applying kainic acid or muscimol to the cortical surface, they eliminated all impulse responses in cortical cells but preserved the impulse activity of the geniculate afferent fibres. When they recorded receptive fields of a number of these fibres, all in the same orientation column, they found that they occupied a long narrow strip that paralleled the orientation of the cells they had previously recorded. Further support came from Tanaka⁸, and from Reid and Alonso⁹, in experiments in which impulses simultaneously recorded from a simple cortical cell and a geniculate cell were cross-correlated. Both studies found that in the two cells, receptive-field regions of the same sign — excitatory or inhibitory — were almost always superimposed.

Finally, two studies from Ferster's laboratory, one published in 1986 and the other in this issue, also point towards excitation from geniculate inputs as the cause of orientation selectivity. Both made use of intracellular recordings of cortical cells, and any reader must surely have wondered why no one had used these methods sooner: intracellular recording allows one to define the properties of a cell in terms of its impulse responses, and then to examine separately the summed excitatory and inhibitory inputs. The answer, of course, is that making such recordings requires consummate skill plus the patience and determination of Job. But if the recordings are successful they can give simple, decisive answers.

In the earlier paper¹⁰, Ferster showed that, in response to visual stimulation, the summed inhibitory inputs to a simple cell, in the form of inhibitory postsynaptic potentials, showed the same orientation preference as the cell itself, with the same peak orientation and identical sharpness of tuning. If the function of the inhibition

were that of sharpening the tuning, the orientation preference of the inhibition should have been different or at least broader.

In the second set of experiments¹, Ferster and his colleagues recorded intracellularly using the whole-cell patch technique, which is roughly equivalent to intracellular recording but more powerful — and still more difficult. They first determined the orientation selectivity of the cell and then suppressed the local impulse activity in that part of the cortex by surface cooling. This should abolish all visually evoked synaptic potentials except those arising from direct geniculate inputs — and indeed it did abolish all the inhibitory input and a fair proportion of excitatory inputs. The receptive fields of the remaining excitation had exactly the same orientation preference and sharpness of tuning as the cell itself before cooling. I find this compelling evidence that the orientation selectivity depends not on inhibitory inputs from neighbouring cortical cells (themselves activated by geniculate inputs, directly or over several stages), or indeed on excitatory polysynaptic inputs that might also serve to sharpen selectivity (by a previously proposed process that others have termed 'selective amplification'). As Ferster *et al.* point out, however, it is still possible that some excitatory input of cortical origin may have survived the cooling; layer 6, the furthest

from the cooling element, is known to send excitatory input to layer 4.

Thirty-five years ago Wiesel and I would have been incredulous had anyone suggested that only now would our scheme for explaining simple cells be vindicated or disproved. At this rate we may expect to have a verdict on a similar proposal we made for complex cells by 2031. Extracellular recordings can tell us about the transformations the cortex makes in the information coming into it, but it is far harder to learn exactly how these transformations come about, in terms of the wiring and the excitatory and inhibitory connections. Ferster and his colleagues have shown that it is possible. □

David Hubel is in the Department of Neurobiology, Harvard Medical School, 220 Longwood Avenue, Boston, Massachusetts 02115, USA.

1. Ferster, D., Chung, S. & Wheat, H. *Nature* **380**, 249–252 (1996).
2. Kuffler, S. W. *J. Neurophysiol.* **16**, 37–68 (1953).
3. Hubel, D. H. *J. Physiol.* **147**, 226–238 (1959).
4. Hubel, D. H. & Wiesel, T. N. *J. Physiol.* **160**, 106–154 (1962).
5. Blakemore, C. & Tobin, E. A. *Expl Brain Res.* **15**, 439–440 (1972).
6. Sillito, A. M. *J. Physiol., Lond.* **250**, 305–329 (1975).
7. Chapman, B., Zahs, K. R. & Stryker, M. P. *J. Neurosci.* **11**, 1347–1358 (1991).
8. Tanaka, K. *J. Neurophysiol.* **49**, 1303–1318 (1983).
9. Reid, R. C. & Alonso, J.-M. *Nature* **378**, 281–284 (1995).
10. Ferster, D. J. *J. Neurosci.* **6**, 1284–1301 (1986).
11. Hubel, D. H. *Eye, Brain, and Vision* (Scientific American Library, New York, 1988).

EVOLUTION

The games lizards play

John Maynard Smith

IN evolutionary theory, an evolutionarily stable strategy (ESS) is a strategy, or phenotype, that is stable in the sense that, if most members of a population adopt that strategy, or have that phenotype, no alternative mutant strategy can invade the population. In simple models, there is usually either a 'pure' ESS (always do the same thing), or a 'mixed' ESS (sometimes do A, sometimes do B, with fixed probabilities). In the latter case, it may also be stable for some individuals always to do A, and others always to do B: that is, there may be a stable genetic polymorphism, instead of individuals adopting a mixed strategy. For some models, however, there is no ESS: the frequencies of different strategies in the population cycle indefinitely. It seems that Sinervo and Lively (page 240 of this issue¹) have now found a real example of that phenomenon.

The logic of the case they describe is that of the 'rock-scissors-paper' game: rock beats scissors, scissors beat paper, and paper beats rock. In the side-blotched lizard, *Uta stansburiana*, males have one of three throat colours, each associated

with a different behaviour. The difference between colour morphs is highly heritable. Orange-throated males establish large territories, within which live several females. A population of such males can be invaded by males with yellow-striped throats: these 'sneaker' males do not defend a territory, but steal copulations. The orange males cannot successfully defend all their females. However, a population of yellow-striped males can be invaded by blue-throated males, which defend territories large enough to hold one female, which they can defend against sneakers. Once sneakers become rare, it pays to defend a large territory with several females. Orange males invade, and we are back where we started from.

The empirical support for these conclusions is as follows. The frequencies of the three morphs were followed for a complete cycle, lasting six years. The fitness of individual males was estimated from the number of females they monopolized, or shared with neighbours. The regression of fitness on the type of neighbour was also estimated. These fitness values were of

precisely the kind to drive the cycle. In particular, no morph was fitter than all others when it was common, so no pure ESS exists.

There is a special pleasure when a curious piece of natural history fits a theoretical prediction. The observation of three distinct male strategies is not new. Gross and Charnov² described three kinds of males in bluegill sunfish: territorial males,

sneakers and female mimics. But it turned out that there are only two lifetime strategies: grow big, and hold a mating territory, or sneak when young, and then, when too big to hide, pretend to be a female and enter the mating territory that way. This is an example of a two-strategy mixed ESS, or, more probably, a genetic polymorphism. In stoats³, there are also three male strategies: territory holders, sneakers

and 'roamers'. However, the strategies again turn out to be in part dependent on age. Young males either hold territories, or sneak copulations. Older males are dominant in interactions, even with territory holders, and roam around acquiring copulations. In neither of these cases is there any reason, in theory or observation, to suspect cycling.

The mere fact that, in the lizard case, no morph is uninvadable is not sufficient to ensure a frequency cycle: there might be a stable polymorphism. Analysis of the rock-scissors-paper game suggests that the cause of cycling, as opposed to polymorphism, may be subtle and hard to measure. However, a simulation incorporating the measured fitnesses of the three lizard morphs did in fact cycle.

Are there other games that cycle? If only two pure strategies are available, cycles will not happen for plausible pay-off values. But if there is a single variable, size for example, that can vary continuously, there may be no ESS. Instead, there can be a sawtooth oscillation: for example, size may increase gradually until a threshold is reached, when the population can be invaded by very much smaller individuals. Myself and Brown⁴ tentatively suggested that such a game might account for the apparently anomalous observation that 90 per cent of mammalian lineages get larger with time, yet the size distribution across species does not increase. We proposed that lineages get larger because of male-male competition, but large size lowers fitness in contexts other than competition, so that large individuals can be replaced by smaller mutants, or, more plausibly, by a smaller species. This would result in a sawtooth oscillation, with gradual increases and sudden decreases in size. Roughgarden⁵ observed that a reverse type of cycle has occurred in Caribbean lizards. On islands, lizards get smaller as an ecological adaptation, but are then replaced by a larger species through interference competition. Both of these cases involve species replacement: an intraspecific cycle would require invasion by a mutant of implausibly large effect.

It seems, then, that Sinervo and Lively have reported the first example of a population cycle caused by frequency-dependent changes in fitness arising from intraspecific interactions. I suspect that it will not be the last. □

John Maynard Smith is in the School of Biological Sciences, University of Sussex, Falmer, Brighton BN1 9QG, UK.

1. Sinervo, B. & Lively, C. M. *Nature* **380**, 240-243 (1996).
2. Gross, M. R. & Charnov, E. L. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6937-6940 (1980).
3. Sandell, M. *Oikos* **47**, 63-70 (1986).
4. Maynard Smith, J. & Brown, R. L. W. *Theor. pop. Biol.* **30**, 166-179 (1986).
5. Roughgarden, J. *Anolis Lizards of the Caribbean* (Oxford Univ. Press, 1995).

EXTRASOLAR PLANETS

Silhouettes of future worlds

A HOPEFUL assumption among many astronomers, science-fiction readers and those who are pessimistic about the human race is that planetary systems are commonplace. Unfortunately there has been little evidence for extrasolar planets until the past few years, with tentative detections of a few massive planets (*Nature* **378**, 355-359; 1995 and **379**, 397-398; 1996) and increasingly strong indications of the existence of protoplanetary systems, in the form of disks of gas and dust around young stars. But now six disks have been imaged directly by the Hubble Space Telescope, and an analysis of their properties by Mark J. McCaughrean and C. Robert O'Dell (to appear in the *Astronomical Journal* in May) hints that more than half of all stars may form planetary systems.

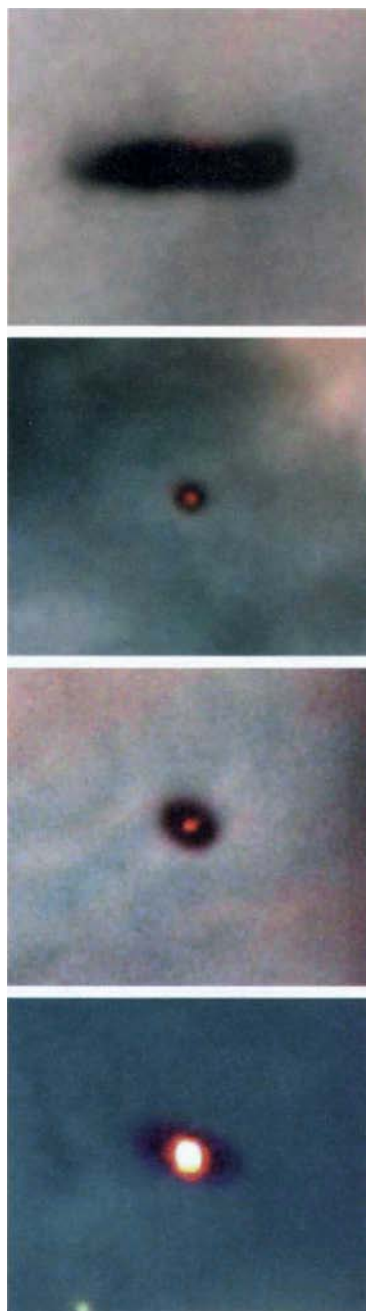
Four of the six are shown here, seen in silhouette against the bright background of the Orion nebula. Their appearances indicate a common disk-like shape seen at different angles, with the uppermost nearly edge-on, hiding its central star. The stars are between 0.8 and 3 million years old and similar to the Sun in size, from 0.3 to 1.5 solar masses.

Contrary to theorists' expectations, McCaughrean and O'Dell find that the disks have sharp edges. Encounters with other stars may be responsible for trimming the disks, in which case the implied rate of encounters is consistent with their being able to survive for about ten million years, which should be long enough to form planets.

The authors also note the significance of similarly dark components in many of the small ionized circumstellar clouds in Orion, and of an infrared 'excess' in the spectrum of 60 to 80 per cent of all the stars in the cluster. It seems that disks are indeed common. But is there enough material in them to make planets?

McCaughrean and O'Dell can only put lower limits on the mass of dust, because it is likely that the disks are opaque at optical wavelengths, and so the space telescope can't 'see inside' them. The lower-limit masses are not enough to form planetary systems like ours. A better estimate can be made using millimetre wavelengths, where the disks are almost transparent. Until those results come in — perhaps within the next few months — we can all be much more justifiably hopeful, if not yet quite convinced.

Stephen Battersby



M. J. McCaughrean, C. R. O'Dell & NASA

Foamy viruses bubble on

Robin A. Weiss

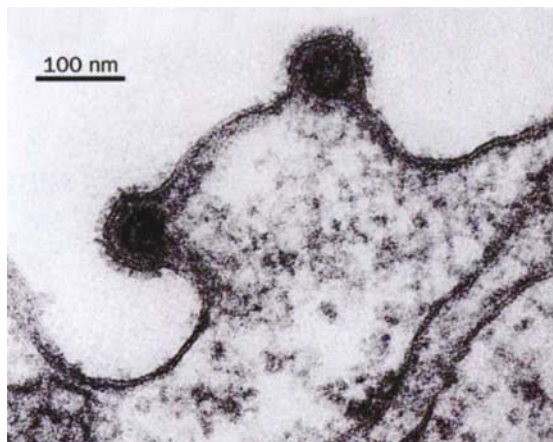
FOR many years we have assumed that all retroviruses replicate by a uniform major strategy. Given that, a paper in last week's *Science*¹ provides pause for thought — it shows that the spumavirus human foamy virus (HFV) differs from oncoviruses (such as mouse leukaemia virus, MLV) and lentiviruses (such as HIV) in carrying full-length, double-stranded DNA genomes in a proportion of virus particles. Furthermore, HFV reverse transcriptase and other Pol proteins are translated from a spliced messenger RNA, and lack Gag sequences^{1,2}. These findings are unexpected and provocative; HFV resembles hepatitis B virus, which has a DNA genome, in these respects as much as it does retroviruses. Moreover, other reports suggest that HFV is not so human after all³, but may serve as a useful gene vector to non-proliferating cells⁴.

Retroviruses have an RNA genome but insert a DNA copy of it into the host chromosome following reverse transcription. The genomes of the three retrovirus subfamilies (onco-, lenti- and spumaviruses) each carry at least three genes, *gag-pol-env*, running 5' to 3', which are packaged in virions as an RNA dimer, with cellular transfer RNA acting as a primer for reverse transcription. It was previously thought that DNA synthesis is only initiated in the cytoplasm of the newly infected cell. Yu *et al.*¹ show that full-length, DNA proviral genomes, some 12 kilobases long, are present in 10–15 per cent of the particles. By contrast, analysis with the polymerase chain reaction indicates that about 1 in 100,000 virions of onco- or lentiviruses have associated full-length DNA. However, the HFV virion DNA does not appear to be infectious and inhibitors of reverse transcriptase indicate that the enzyme is still required at an early stage of infection.

Although foamy viruses have essentially the same main gene organization as other retroviruses (plus accessory *bel* genes which lie 3' to *env*), the finding that *pol* is translated from its own spliced mRNA, independently of *gag*, is also quite novel. Other retroviruses synthesize Pol products as part of a Gag–Pol fusion protein, translated either by ribosomal –1 frameshift towards the end of Gag, or by use of a suppressor tRNA. Thus Pol is incorporated into virions via Gag domains in the precursor protein before cleavage by viral protease. As Yu *et al.* point out, their results pose the question of how Pol is assembled into viral particles.

When I last wrote about foamy viruses in these columns⁵ I called them viruses in search of a disease. Although they are the most cytopathic, syncytium-inducing retroviruses for many cell types *in vitro*, disease is seldom seen in the many species of pri-

mate that they naturally infect. However, mice transgenic for HFV suffer neurodegeneration resembling amyotrophic lateral sclerosis⁶, and virus expression was evident in the brain of an orang-utan which died with similar acute neurological symptoms⁷. But HFV has not been found in human amyotrophic lateral sclerosis⁸, and the various claims that HFV is associated with de Quervain⁵ or Graves⁹ diseases of the thyroid has now been questioned¹⁰. Indeed there is no longer definitive evidence of human infection, other than in a handful of



Budding particles of foamy retrovirus, imaged by electron microscopy; note the prominent glycoprotein spikes and the preformed spherical cores. As seen through the light microscope, cell cultures infected with the virus look foamy, hence the name — spumaviruses — of the subfamily.

persons with zoonosis from contact with primates³. HFV was the first 'human' retrovirus to be described¹¹, yet it is now apparent that the HFV genome sequence is indistinguishable from that of one of the chimpanzee isolates¹². Following an international workshop on foamy viruses in 1994, at which all laboratories working on foamy virus were represented, exchange and analysis of samples did not uphold evidence for human foamy virus in patients with autoimmune disease.

Given that the great apes so frequently harbour foamy viruses^{7,12,13}, it is odd that *Homo sapiens* has escaped endemic infection. The few persons who have become infected with simian foamy virus (SFV) remain well, have a 'full house' of antibody responses analysed by Western blot, and have viral genomes easily detected in blood by the polymerase chain reaction³. These observations have two implications: that reports of widespread human infection are tenuous at most; but that humans can nonetheless be infected by SFV, and in all probability would become infected following xenotransplantation of primate tissue. Thus the patient with AIDS who has received a baboon bone-marrow

transplant, and any future organ transplants from primates, should be carefully monitored for foamy virus transfer as well as other viruses.

Parodying Koch's postulates for modern virology, one might propose that if a virus fails as a pathogen, promote it as a vector in gene therapy. This is precisely what has been done for HFV. Schmidt and Rethwilm¹⁴ deleted the *bel2-3* region, while Russell and Miller⁴ substituted marker genes in place of *env* and *bel1*. With its very broad host range and tropism for cell types⁴, HFV may prove to be quite a useful vector.

Russell and Miller further noted that their foamy virus vectors were able to infect cells in the stationary phase of the cell cycle, and might thus be used to deliver genes to non-proliferating targets such as neurons. On the face of it, this finding appears to run counter to our recent report¹⁵ that HFV replication is cell-cycle dependent. These apparently contradictory results stem largely from a different emphasis in interpretation, however — in both cases, the efficiency of infection of stationary cells was approximately 5–10 per cent that of proliferating cells. Thus HFV infection is less stringently dependent on mitosis than MLV, but is more dependent than HIV. An attractive explanation for HFV infection of non-proliferating cells would be that it occurs through the DNA genomes found in 10–15

per cent of virions (though Yu *et al.*¹ thought that the DNA was not infectious). Perhaps Russell and Miller should check whether their vectors can infect stationary cells in the presence of inhibitors of reverse transcriptase. □

Robin A. Weiss is in the Chester Beatty Laboratories, Institute of Cancer Research, 237 Fulham Road, London SW3 6JB, UK.

1. Yu, S. F., Baldwin, D. N., Gwynn, S. R., Yendapilli, S. & Linial, M. L. *Science* **271**, 1579–1582 (1996).
2. Enssle, J., Jordan, I., Mauer, B. & Rethwilm, A. *Proc. natn. Acad. Sci. U.S.A.* (in the press).
3. Schweizer, M. *et al. AIDS Res. hum. Retrovir.* **11**, 161–170 (1995).
4. Russell, D. W. & Miller, A. D. *J. Virol.* **70**, 217–222 (1996).
5. Weiss, R. A. *Nature* **333**, 497–498 (1988).
6. Bothe, K., Aguzzi, A., Lassman, H., Rethwilm, A. & Horak, I. *Science* **253**, 555–557 (1991).
7. McClure, M. O. *et al. J. Virol.* **68**, 7124–7130 (1994).
8. Rösener, M., Kranz, M., Hahn, H., Heeney, J. & Rethwilm, A. *J. med. Virol.* (in the press).
9. Lagaye, S. *et al. Proc. natn. Acad. Sci. U.S.A.* **89**, 10070–10074 (1992).
10. Schweizer, M., Turek, R., Reinhardt, M. & Neumann-Haefelin, D. *AIDS Res. hum. Retrovir.* **10**, 601–605 (1994).
11. Achong, B. G., Mansell, P. W. A., Epstein, M. A. & Clifford, P. J. *J. gen. Virol.* **40**, 175–181 (1971).
12. Herchenröder, O. *et al. Virology* **201**, 187–199 (1994).
13. Bieniasz, P. D. *et al. Virology* **207**, 217–228 (1995).
14. Schmidt, M. & Rethwilm, A. *Virology* **210**, 167–178 (1995).
15. Bieniasz, P. D., Weiss, R. A. & McClure, M. O. *J. Virol.* **69**, 7295–7299 (1995).

The brine of the Earth

Terry Plank

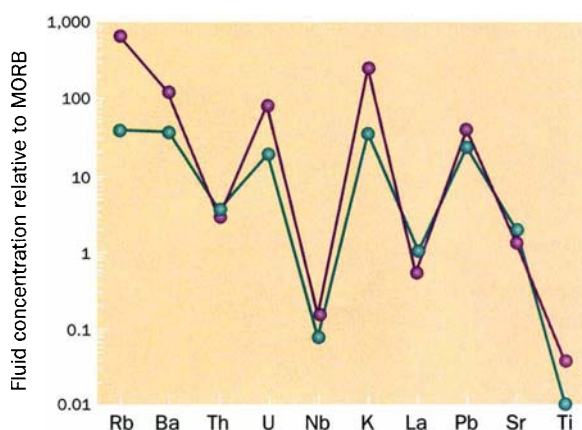
ONE of the most remarkable processes occurring in the Earth is the circulation of sea water down subduction zones and back out of 'arc' volcanoes. Sea water is bound in hydrous minerals until the great pressure of subduction causes them to break down, letting the released water rise to the crust, dissolved in erupting magmas. This cycle is largely inferred from the wet crustal material we see entering the subduction zone at oceanic trenches and from the volatile-charged eruptions from arc volcanoes. The path in between, through the mantle, is unseen and still largely unknown. The reactions that bind and later release water from the downgoing slab, and the actual nature of the fluid phase — whether a hydrous silicate melt, a silica-rich hydrous fluid, or a saline fluid — are still debated. On page 237 of this issue, Keppler¹ provides fresh evidence that salty water is the furtive fluid phase that accounts both for the melting that leads to arc volcanism and for the transport of trace elements that have long characterized the geochemistry of arc magmas.

Volcanic rocks from convergent plate margins possess a unique geochemical fingerprint, enriched in alkalis and alkaline elements (rubidium and barium for example) and depleted in highly charged cations (such as niobium and titanium). Based mostly on the behaviour of these elements in low-temperature water-rock reactions, geochemists have guessed that the mechanism responsible for the chemical 'arc fingerprint' is dehydration of the downgoing slab, and element-selective partitioning into the released fluid. Most models end here, because few experiments have been conducted to determine the partitioning characteristics of fluids at appropriate pressures and temperatures.

That is not for lack of interest or trying, but because such experiments are technically challenging. There are problems with keeping the fluid from leaking out of the experimental capsule, allowing enough time for fluid and solid to reach equilibrium, and retrieving the fluid to analyse it for trace elements. Keppler's approach¹ avoids some of the problems of fluid-solid equilibration by leaving the solid out. By measuring $D^{\text{fluid/melt}}$, the ratio of the fluid concentration to the melt concentration for a given element, the known value of $D^{\text{solid/melt}}$ can be used to obtain $D^{\text{fluid/solid}}$ indirectly.

Keppler focuses on results obtained for

highly saline fluids and the mineral clinopyroxene at 3 kbar and 1,040 °C. Because clinopyroxene will be an important mineral phase in the high pressure slab (as eclogite) and in the mantle, these experimental results are clearly relevant to slab dehydration, as well as fluid migration in the mantle². The fluid used, a five molal sodium chloride/potassium chloride mixture, leads to $D^{\text{fluid/solid}}$ values close to those inferred for arc magmas, with pref-



Trace-element compositions for fluids in equilibrium with mid-ocean-ridge basalt (MORB). The results for a five molal NaCl-KCl fluid¹ (aquamarine) are compared with those for altered MORB (ref. 6.) and lower salinity, using an average between Keppler's 5 and 0 molal NaCl results (purple). Both are calculated assuming 2.5 per cent fluid fraction. The two patterns are very similar to one another, and to the enrichments observed in arc magmas and the continental crust, so subduction fluids may not need to be extremely saline.

erential partitioning of rubidium, barium and lead into the fluid, and unmeasurable quantities of niobium and zirconium.

The experiments also provide a way to explain other features of arc magmas. The low cerium/lead ratio³ of arc magmas has been used to argue for a fluid phase that extracts lead from the subducted oceanic crust more effectively than it does cerium, and Keppler's five molal NaCl fluid does just that — by almost two orders of magnitude. Another observation that provides evidence for fluid partitioning in subduction zones is the $^{238}\text{U}/^{230}\text{Th}$ activity ratio, which is greater than what would be expected from radioactive equilibrium⁴. Again, this is supported by the laboratory saline fluids, which produce $D^{\text{fluid/solid}}$ for uranium ten times higher than that for thorium.

There is a broader implication to these results. The upper continental crust shares many geochemical features with arc magmas, supporting the idea that continents grow by accreting arc magmas. Taken at face value, Keppler's experiments lead to the conclusion that the geochemical make-up of the continental crust is the

work of a single substance, salty water.

Do fluids have to be so salty to bring about the right partitioning? (Five molal NaCl is about ten times saltier than sea water.) Because subducting slabs are already enriched in many of the same trace elements as arc lavas⁵, fluid partitioning effects might not have to be so large. It is possible to recreate virtually the same trace-element pattern as the five molal NaCl fluid, by using partitioning ratios corresponding to an intermediate salt concentration and starting with altered oceanic basalt⁶ instead of pristine mid-ocean-ridge basalt (see figure). More dramatic enrichments, requiring even lower salinity, may be obtained if subducted sediments are included, so extreme salinity may not be a requirement of successful arc geochemistry models.

There is also more than one way to squeeze water out of a rock. Using a more direct experimental approach, Brenan *et al.*⁷ determined fluid/mineral D s for water rich in other solutes — silicon, aluminium and magnesium. Even without chlorine, Brenan and co-workers' experiments can reproduce many of the geochemical features of arc basalts, such as their low cerium/lead ratio⁸. This convergence of experimental results provides compelling support for an aqueous fluid phase in the source of arc magmas.

Nonetheless, there are some clear differences between the results. In the absence of highly saline fluids, Brenan *et al.* do not find a very low $D^{\text{fluid/solid}}$ for niobium, but point out the importance of the mineral rutile in retaining niobium in the solid slab⁹. Other residual minerals such as mica, serpentine and amphibole will also have a clear effect on element partitioning. So the results of Brenan *et al.* raise an important point: other solutes besides NaCl and other minerals besides clinopyroxene are possible players.

A perplexing twist in the salinity story comes from experiments¹⁰ that do not show big differences in partitioning between

1. Keppler, H. *Nature* **380**, 237–240 (1996).
2. Stolper, E. M. & Newman, S. *Earth planet. Sci. Lett.* **121**, 293–325 (1994).
3. Miller, D. M., Goldstein, S. L. & Langmuir, C. H. *Nature* **368**, 514–520 (1994).
4. Gill, J. B. & Williams, R. W. *Geochim. cosmochim. Acta* **54**, 1427–1442 (1990).
5. Plank, T. & Langmuir, C. H. *Nature* **362**, 739–742 (1993).
6. Staudigel, H., Plank, T., White, W. M. & Schmincke, H.-U. in *Dynamics of Subduction* (eds Bebout, G. E., Scholl, D. & Kirby, S.) (AGU Monogr., Washington DC, in the press).
7. Brenan, J. M., Shaw, H. F., Ryerson, F. J. & Phinney, D. L. *Geochim. cosmochim. Acta* **59**, 3331–3350 (1995).
8. Brenan, J. M., Shaw, H. F. & Ryerson, F. J. *Nature* **378**, 54–56 (1995).
9. Brenan, J. M., Shaw, H. F., Phinney, D. L. & Ryerson, F. J. *Earth planet. Sci. Lett.* **128**, 327–339 (1994).
10. Ayers, J. C. & Eggler, D. H. *Geochim. cosmochim. Acta* **59**, 4237–4246 (1995).
11. Peacock, S. M. *Chem. Geol.* **108**, 49–59 (1993).

melt and saline fluid. At high pressures (>15 kbar) the main effect of adding chlorine is simply to increase the solubility of silicate melt in fluid. Perhaps it is not chloride complexes but total solute content that controls element partitioning. Whether the differences between the two salt-water studies are due to the different pressure ranges explored or to experimental or analytical problems remains to be tested, but linking element partitioning to total solute content provides one explanation for some of the similarities between the saline and salt-free results.

A final word of caution — all of these experiments were conducted at 900 °C or more. Most slab dehydration is thought to occur at much lower temperatures¹¹, where silicate solubility and fluid par-

tioning may be considerably depressed. Also, at slab temperatures greater than 900 °C silicate melting might become important.

This sudden wealth of fluid/crystal partitioning data has provided welcome laboratory confirmation of the role of recycled sea water in the creation of arc magmas and the continental crust. Many issues remain outstanding, in particular the role of different fluid solutes. Studies of arc magmas where chlorine and water are combined with trace elements may be needed to identify the importance of salty water in subduction recycling. □

Terry Plank is in the Department of Geology, University of Kansas, Lawrence, Kansas 66045, USA.

HAEMOGLOBIN

Root cause of fish buoyancy

Rory Howlett

FISH have a problem — because their body tissues are denser than water they have a tendency to sink. Those that have solved the problem with a swim bladder face the complication that, to retain neutral buoyancy, they must secrete gas into the bladder against a pressure gradient; the deeper they swim, the bigger the gradient. A study by Mylvaganam *et al.*, published in the March issue of *Nature Structural Biology*¹, in part reveals how they do it. The authors show that certain fish haemoglobins have structural characteristics that make them efficient proton-driven oxygen pumps, able to drive the gas into a swim bladder.

Cartilaginous and bony fish retain buoyancy in different ways. Cartilaginous fish such as sharks never cease swimming, using their pectoral fins as hydrofoils and their asymmetrically shaped tail fins to generate lift, and also accumulate low-density oils and the hydrocarbon squalene in their tissues. Most pelagic and surface-living bony fish have a swim bladder which is essentially an extensible sack of gases, mostly nitrogen and oxygen, and which acts as a kind of portable float.

Gas has clear advantages over squalene, but its use in an aqueous environment also presents challenges. As a fish swims deeper it experiences higher water pressure and its swim bladder is compressed, while the opposite occurs as the fish moves towards the surface. To compensate for these changes, as well as diffusional losses, the fish has continually to regulate the amount of gas in the swim bladder, and here things can become tricky. Some species replenish gas supplies simply by gulping air at the water surface which is then passed through a duct to the swim bladder. But others, such as those

living in deeper water or lacking a bladder duct, must secrete gases from the blood to the swim bladder against a pressure gradient. This is generally achieved through the ingenious design of the so-called gas gland, which secretes lactic acid. One effect of the acid is to force nitrogen out of solution, but the most striking consequence is a marked reduction in the oxy-

induced conformational changes in fish haemoglobins, although the precise mechanism has remained unclear.

To address this issue, Mylvaganam and colleagues studied the structure and function of a representative Root-effect haemoglobin from *Leiostomus xanthurus*, commonly known as the spot, a bony fish that lives along the eastern seaboard of the United States. Using X-ray crystallography, they have determined the structure of the ligand-bound protein at 2 Å resolution, and have identified several key amino-acid residues that assemble into two positive-charge clusters across the interface of two β -chains in the high-affinity R-state haemoglobin. Binding of a proton flips an electrostatically based switch, pushing apart the two positive-charge clusters, and inducing a profound conformational change. This effectively locks the Root-effect haemoglobin in the low-affinity T-state even at high pressures of pure oxygen, which would account for the remarkable effectiveness of Root-effect haemoglobins as proton-driven oxygen pumps.

Aspects of the proposed mechanism will have to be confirmed by mutagenesis studies, to assess the effect of specific amino-acid substitutions in the two positive-charge clusters. Meantime, comparisons of the sequences of various fish haemoglobins should provide pointers as to how the magnitude of the Root effect has been fine-tuned by natural selection to suit the specific needs of fish with different life histories and ecology. More generally, the new work opens the way to the rational engineering of haemoglobins with defined oxygen-binding characteristics that could be of importance in medicine and elsewhere.

This is just one of the latest of a series of findings unveiling the marvellous properties of haemoglobin: one only has to turn the page to find discussion of another. And last year, for instance, Komiyama *et al.*² showed that the allosteric

Spot the difference — blue marlin, like the subject of Mylvaganam and colleagues' studies, has Root-effect haemoglobin.

gen affinity of blood haemoglobin — a phenomenon known as the Root effect.

It is this pH-dependent release of oxygen from Root-effect haemoglobin, as well as some fancy plumbing in the gas gland itself, that largely accounts for the ability of fish to compress oxygen and force it into the swim bladder when extra buoyancy is needed. But exactly why is the oxygen affinity of fish haemoglobin so exquisitely sensitive to pH? The oxygen binding and release characteristics of haemoglobins involve allosteric changes, and it seems reasonable to assume that the Root effect is driven by proton-

characteristics of crocodile haemoglobin are determined by a few key amino-acid residues; these give rise to the adaptive oxygen-binding characteristics of haemoglobin suited to the respiratory needs of crocodiles, which often remain submerged for long periods when drowning large prey. □

Rory Howlett is Deputy Biological Sciences Editor of *Nature*.

1. Mylvaganam, S. E., Bonaventura, C., Bonaventura, J. & Getzoff, E. D. *Nature struct. Biol.* **3**, 275–283 (1996).
2. Komiyama, N. H., Miyazaki, G., Tame, J. & Nagai, K. *Nature* **373**, 244–246 (1995).

The real water-bearer

Sun Kwok

THE Infrared Space Observatory (ISO), launched by ESA last November, has detected water in the planetary nebula NGC7027. Given that water is widely present on Earth, it would seem that it is a natural molecule to be found in the interstellar medium. In fact, the ISO discovery is the first success after many years of searching, and a major step in solving a long-standing puzzle of interstellar chemistry — the missing oxygen problem.

It is generally accepted from solar observations that oxygen is the third most abundant element (after hydrogen and helium) in the Universe. In fact there is believed to be more oxygen than all the other elements combined. It has been detected in various forms in the interstellar medium — the strong green lines in the spectra of gaseous nebulae are from doubly-ionized oxygen, and infrared spectroscopy has seen oxygen in solid form (as ice, for example, or as part of the sand-like silicate compounds commonly seen around cool stars and dark clouds). It is also seen in molecules such as carbon monoxide, which has been widely detected in the interstellar medium. However, only about half of the oxygen implied by solar abundances is accounted for, and the rest remains 'missing'.

Theoretical chemical models imply that most of the oxygen should be in the form of water and O_2 molecules, but the detection of these molecules has proved elusive. Because of the pressure-broadened absorption lines of the same molecules in Earth's atmosphere, most of the line emissions from celestial sources don't reach the ground. Balloon experiments have searched for oxygen, but so far without success.

Although water has been detected in various lines from natural masers, reliable abundances cannot be derived from these observations¹. The water molecule, being asymmetric, has a complicated spectrum. It is difficult to translate the maser intensity arising from a highly excited state into a total molecular abundance. The most straightforward way to determine abundance is to measure radiation from the ground-state transition.

Attempts have been made to find molecules containing different isotopes of oxygen and hydrogen, which would radiate at frequencies less affected by atmospheric absorption. Astronomers have also tried to use the fact that molecular lines from distant galaxies are redshifted, and so could pass through spectral windows in the atmosphere. All results to date have been negative^{2,3}.

The most extensive effort has been made by the Cosmic Background Explorer satellite (COBE), which carried a spec-

trometer capable of measurements in the 100 μm to 1 cm wavelength range. There are a number of important atomic and molecular transitions in this range that are difficult to observe from the ground. With a wide 7-degree beam, COBE was successful in mapping the Galactic plane, and detected emission lines from carbon and nitrogen. COBE found that the 157 μm line from Galactic C II (singly ionized carbon) radiates as much energy alone as 100 million Suns⁴.

This radiation escapes into the intergalactic medium, so atomic and molecular lines are powerful agents in the cooling of the Galaxy. However, COBE was only able to put upper limits on water and oxygen abundances. As they are not widely distributed in the Galactic plane, they must be much less abundant than we thought, or they are only concentrated in small volumes. COBE, with its wide beam, will fail to detect signals from small sources. This has brought home the need for a large antenna (therefore a narrow beam) to search for possible locations of oxygen and water molecules, such as stellar envelopes or cores of star-forming regions.

The ISO results are not only a positive step in the solution of the missing oxygen mystery, they also pose some new questions. The object in which water was

detected (NGC7027) is a very young planetary nebula with a massive envelope of dust and molecular gas left over from its progenitor, a giant star whose stellar wind formed the nebula. Many molecular species have been seen in this object. However, it is carbon rich, and in traditional chemical models most of the oxygen gets tied up in CO in such an environment, and not much water is expected. This observational discovery therefore poses new challenges to our understanding of interstellar chemistry, and a new network of reactions may have to be developed to account for this surprising result.

The water transition detected by ISO is the 179.5 μm rotational line. When the flux of this line has been absolutely calibrated, reliable abundances of water can be derived. A different line at 539 μm from the ground-state transition of water is the target of two upcoming space missions, SWAS and Odin. The ISO results have raised high hopes that these missions will bring in enough new data finally to identify the whereabouts of the Universe's oxygen. □

Sun Kwok is in the Department of Physics and Astronomy, University of Calgary, Calgary, Canada T2N 1N4.

1. Menten, K. M., Melnick, G. J. & Phillips, T. G. *Astrophys. J.* **350**, L41–L44 (1990).
2. Combes, F., Casoli, F., Encarnaz, P., Gerin, M. & Laurent, C. *Astr. Astrophys.* **248**, 607–612 (1991).
3. Liszt, H. S. *Astrophys. J.* **386**, 139–142 (1992).
4. Wright, E. L. et al. *Astrophys. J.* **381**, 200–209 (1991).

BLOOD

Taking the pressure off

M. F. Perutz

THE identification of the endothelial relaxing factor with nitric oxide (NO^*), an unstable and highly toxic radical, has been one of the most astonishing discoveries in physiology^{1–3}. The ramifications have been seemingly endless — only last week, for instance, this journal carried a report⁴ of endothelial damage caused by superoxide radicals generated by β -amyloid, which inactivate nitric oxide.

One remaining puzzle is that the haem iron in haemoglobin has an affinity for NO^* at least ten thousand times greater than for oxygen. In the walls of our blood vessels, NO^* is generated from arginine by the haem-enzyme nitric oxide synthase³. Why is NO^* liberated there not instantly sequestered and inactivated by the erythrocytes?

On page 221 of this issue⁵ Stamler and his colleagues demonstrate the existence of an equilibrium between erythrocytic NO^* bound to the haem iron and the reactive thiol groups of haemoglobin on the one hand, and NO^* bound to glu-

tathione and other small molecular erythrocytic thiols on the other hand. They also show that these thiols can transfer NO^* from erythrocytes to endothelial receptors and relax vascular tension.

This work is the continuation of research made possible by the use of a highly sensitive chemiluminescence method for the quantitative determination of nanomolar concentrations of nitrosothiols in biological systems⁶, in a prolonged study of the interaction of NO^* with thiol groups in peptides and proteins *in vitro* and *in vivo*. The results showed that glutathione is abundant in human lung alveoli; that *in vivo* NO^* reacts with it, forming S-nitroso (SNO)-glutathione which has a half-life of several hours compared to that of a few seconds of NO^* *in vivo*; and that SNO-glutathione reduces the toxicity of NO^* and relaxes smooth muscle in human airways. The studies also showed that SNO-cysteine is a more potent vasorelaxant than SNO-glutathione and more potent even than the classical relaxant nitroglycerine, probably because it

decomposes rapidly into cysteine and NO[•]; the authors also suggested that some of the NO[•] generated from arginine at the haem iron of nitric oxide synthase may be carried to its receptors by thiols⁷⁻¹¹, but there is no convincing evidence for this¹².

The new results⁵ show that, in blood, NO[•] is bound to both the haem iron and the reactive sulphhydryl (SH) groups of cysteine 93 β of haemoglobin. (In arterial blood [FeNO] = 536 $\pm$ 99 nM and [cysteine 93 β SNO] = 311 $\pm$ 55 nM. In venous blood [FeNO] = 900 $\pm$ 126 nM and [cysteine 93 β SNO] is negligible). The concentration of haem iron in blood is about 9 mM, which makes the concentration of NO[•] carried by the blood about 1/10,000 of that of oxygen. NO[•] bound to haemoglobin in the absence of oxygen is relatively stable, but in its presence NO[•] is instantly converted to nitrate and the haem iron is oxidized to methaemoglobin which is then reduced by methaemoglobin reductase.

Reaction of NO[•] with the SH groups of cysteine 93 β is controlled by the allosteric transition of haemoglobin and by the spin state of the haem iron: reactivity is high in the oxy or R-structure and low in the deoxy or T-structure¹³ and, within the R-structure, the off-rate of NO[•] is faster in high-spin methaemoglobin than in low-spin oxyhaemoglobin. Nature exploits this regulation. Liberation of oxygen in the tissues, accompanied by the allosteric transition from R to T, or by oxidation of the haem iron, causes NO[•] to be liberated by cysteine 93 β and to be transferred to small thiols which may return it to endothelial receptors.

The authors proved their point by preincubating human erythrocytes with nitrosocysteine and then washing off all residual nitrosocysteine. The S-nitroso-oxyhaemoglobin thus produced in the erythrocytes transferred its NO[•] to small thiols which were exported to the endothelium. On the other hand, the authors report that treatment with dithionite, which removed the oxygen and converted haemoglobin to the T-structure, caused NO[•] to be transferred from the SH groups of cysteine 93 β to the haem irons, but this seems unlikely because dithionite instantly destroys NO[•]. Incubation of oxy- or deoxyhaemoglobin with either SNO-glutathione or SNO-cysteine resulted in transfer of NO[•] to cysteine 93 β , but in hardly any transfer of NO[•] to the haem irons.

By sequestering endothelial NO[•], solutions of free haemoglobin caused vasoconstriction both *in vitro* and *in vivo*. But when its SH groups were blocked with N-ethylmaleimide, the rise in tension was 20 per cent less; when the haem irons were oxidized and blocked with cyanide, it was 70 per cent less. This proved that both the haem irons and the SH groups act as NO[•] scavengers. Nitrosylation of the SH group inhibited the rise in vascular tension produced by free oxyhaemoglobins. Because

of its low affinity for NO[•], S-nitroso-methaemoglobin actually relaxed tension by release of NO[•]. In the presence of glutathione, S-nitroso-oxyhaemoglobin as well as S-nitrosomethaemoglobin relaxed tension, showing again that NO[•] is transferred to small thiols and then conveyed to NO[•]-receptors. *In vivo*, the methaemoglobin formed in the reaction of NO[•] with the haem iron of oxyhaemoglobin is instantly reduced to deoxyhaemoglobin by methaemoglobin reductase. On reduction of its haem irons, any NO[•] carried by its reactive cysteines would be liberated and could be transferred to small molecular thiols. These observations would explain why erythrocytes, which are rich in glutathione, cause no rise in vascular tension, whereas free haemoglobin dissolved in the plasma, which lacks glutathione, does raise tension. The biosynthetic pathway for the formation of nitrosothiols is not clear, because NO[•] does not nitrosylate thiol directly, but only in the presence of higher nitrogen oxides¹².

In the course of evolution of haemoglobin, only the haem-linked histidine and a phenylalanine which wedges the haem into its pocket have remained invariant; on the other hand, cysteine 93 β has remained invariant in mammals and birds, which implies that it fulfils an essential function, but its nature has remained a mystery. The paper by Stamler and colleagues⁵ now suggests that it acts as an allosterically and iron-spin-controlled nitric oxide buffer, exchanging nitric oxide with low-molecular-weight thiols.

Several firms are engaged in clinical trials of genetically or chemically modified haemoglobins as blood substitutes. In phase I trials on healthy volunteers, infusion of a haemoglobin with covalently linked α -chains¹⁴ caused a transient rise in diastolic pressure of 10–12 mm Hg. This is not serious, but the authors suggest that it could be avoided by the addition of S-nitrosylated haemoglobin. This is an intriguing possibility. □

M. F. Perutz is at the MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

1. Furchgott, R. F. & Zawadzki, J. V. *Nature* **288**, 373–376 (1980).
2. Ignarro, J. L., Byrns, R. E., Buga, G. M. & Wood, K. S. *Fed. Proc.* **46**, 644 (1987).
3. Palmer, R. M. J., Ashton, D. S. & Moncada, S. *Nature* **333**, 664–666 (1987).
4. Thomas, T. et al. *Nature* **380**, 168–171 (1996).
5. Jia, L., Bonaventura, C., Bonaventura, J. & Stamler, J. S. *Nature* **380**, 221–226 (1996).
6. Downes, M. J. et al. *Analyst* **101**, 242–248 (1976).
7. Stamler, J. S. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7674–7677 (1992).
8. Gaston, B. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 10957–10961 (1993).
9. Stamler, J. S. *Curr. Topics Microbiol. Immun.* **196**, 19–33 (1995).
10. Stamler, J. S. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 444–448 (1992).
11. Stamler, J. S. *Cell* **78**, 931–936 (1994).
12. Butler, A. R., Flitney, F. W. & Wilkins, D. L. H. *Trends Pharmacol. Sci.* **16**, 18–22 (1995).
13. Benesch, R. E. & Benesch, R. *Biochemistry* **1**, 735–739 (1962).
14. Looker, D. et al. *Nature* **356**, 258–260 (1992).

Grow your own

WHY are addictive drugs illegal? Society cannot permit too many of its citizens to make themselves useless; but the true emotional reason is subtler. The addict seems to get pleasure directly. He doesn't have to work, or sell things, or jump through any approved social hoops to get it. He no longer needs society; he only needs his drug. This defiance of the rest of us is intolerable. He can't be allowed to get away with it.

The reality, of course, is different. The addict's pleasure gets ever feebler and briefer, and soon he is driven merely by the need to avoid pain — the agony of withdrawal. His drug is illegal; he has no approved way of getting it; so he turns to crime, thus compounding his own miseries and adding to everyone else's. There must be a better way.

In this connection, Daedalus recalls the hope and hype of gene therapy. A few living cells are taken from the patient's body; a new gene is inserted into them by standard methods; and they are then put back. Thereafter, they release into his body the biochemical expressed by the gene. With luck it cures his disease.

The major addictive drugs are, of course, made by genes — those of the opium poppy and the coca bush. Animal and plant genes are quite compatible; firefly genes continue to work inside tobacco plants, which then glow in the dark. So Daedalus hopes to isolate the genes for the main plant narcotics, and find ways of inserting them into human cells. When put back into the donor, the cells will leak a steady supply of the narcotic into his bloodstream. Brain cells, though hard to extract and replace, would be ideal. With such precise targeting, even tiny traces of drug, from very few implanted cells, would be effective.

With a secure supply of his drug inside him, the addict will be freed from his major problem. The pangs of withdrawal will no longer force him into crime. He will not even be able to sell his supply to other addicts. His implant will give him no sinful pleasure — that stage of his addiction passed long ago — but he will at least feel fairly normal. He should be able to shoulder the burdens of responsible citizenship once more.

Such a simple, permanent solution will not be popular. The pushers and drug barons will lose their profits, the addicts will lose their dramatic pose as defiant losers, the police and the drug-clinics will lose a rich field of activity, and the law-abiding citizens will lose the spectacle of wickedness meeting its just deserts. But, says Daedalus, never do by law what you can do by engineering, particularly if the law doesn't work. It seems worth a try.

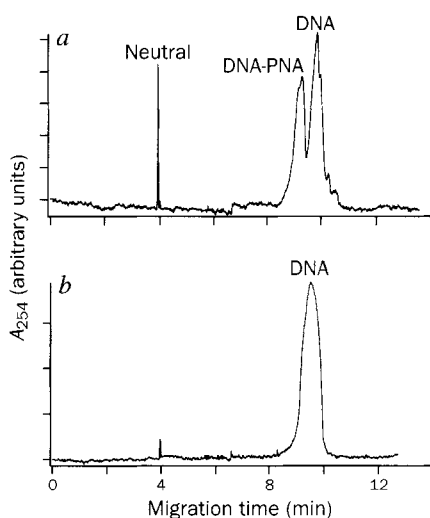
David Jones

Screening for genetic mutations

SIR — A practical method for screening human genes for mutations must be rapid, inexpensive and have the capability of detecting single base substitutions. We report here the use of the DNA mimic, peptide nucleic acid (PNA)¹, as a probe for genetic mutations.

PNA differs from DNA in that the DNA ribose-phosphate backbone is replaced by its pseudo-peptide counterpart, in this case *N*-(2-aminoethyl)glycine. By choosing the sequence of the four bases in PNA to be complementary to one of the strands of DNA, the PNA probe hybridizes by Watson-Crick pairing. Such hybridization reactions can be rapid, and the hybridization products can be readily separated by free-solution capillary electrophoresis because the hybrid complex has a different ratio of electrical and frictional forces from the native DNA, because of the non-ionic nature of PNA². We illustrate this procedure by using a PNA to distinguish normal (wild-type) and mutant sequences in the human cystic fibrosis gene.

The table gives the sequence of the



Comparison of electropherograms at 70 °C of wild-type gene hybridized with PNA (a) and DNA containing a single-base substitution (MU1) in the target region hybridized with PNA (b). Free PNA is not detected because it binds to the capillary wall.

15-mer PNA probe we synthesized, as well as four test sequences, each comprising a 50-mer single-stranded DNA fragment, representing the wild-type and three mutant sequences of the cystic fibrosis transmembrane conduction regulator (CFTR) gene³. In particular, the three-base deletion Δ F508 accounts for two-thirds of the CF mutations, the percentage varying with different populations⁴.

The hybridizations were carried out for 5–10 min at room temperature, and the mixtures were then separated on a Beckman P/ACE Model 2000 capillary electrophoresis instrument (kindly made available by S. Pentoney Jr) modified to operate at high temperatures. The figure shows electropherograms for wild-type (a) and mutant MU1 (b) DNA where the latter is a single base substitution in the target region hybridized with PNA (see table), both recorded in absorption at 70 °C. At this temperature and ionic strength (50 mM) DNA-DNA hybridization is unlikely, and DNA-PNA hybrid complexes melt unless the sequence is a perfect match. The peaks in the figure were identified by varying the PNA concentration in the hybridization reaction (not shown). This ability to distinguish mismatches has been demonstrated for other sequences (not shown).

In other experiments, we added the PNA probe to a polymerase chain reaction preparation of human DNA of the 142-mer wild-type CF fragments and the 139-mer three-base deletion fragments. At high temperature (50 °C), a PNA complex of unknown structure is indicated with the wild-type but not with the mutant DNA. Before the method can be developed to its full potential, there must be further testing of genomic double-stranded DNA and exploitation of the concentration dependence of the PNA-DNA signal to allow distinction between homo- and heterozygosity.

Our results to date encourage us to believe that a PNA probe library, possibly with multiple fluorescence tags for multiplex testing of one or more exons, might

form the basis for a universal screening strategy for any genetic disease with a known spectrum of mutations⁵.

Christina Carlsson, Mats Jonsson* Bengt Nordén

Department of Physical Chemistry,
Chalmers University of Technology,
S-412 96 Gothenburg, Sweden

Maria T. Dulay, Richard N. Zare

Department of Chemistry,
Stanford University,
Stanford, California 94305-5080, USA

Jaan Noolandi

Xerox Research Centre of Canada,
Mississauga, Ontario L5K2L1, Canada

Peter E. Nielsen

Department of Biochemistry B,
Panum Institute,
DK-2200 Copenhagen N, Denmark
Lap-Chee Tsui, Julian Zielenski
Department of Genetics,
The Hospital for Sick Children and
University of Toronto,
Toronto, Ontario M5G 1X8, Canada

1. Egholm, M. *et al.* *Nature* **365**, 566–568 (1993).
2. Noolandi, J. *Electrophoresis* **14**, 680–681 (1993).
3. Welsh, M. J. *et al.* in *Molecular and Metabolic Basis of Inherited Disease* 7th edn (eds Scriver, C. *et al.*) 3799–3876 (McGraw-Hill, New York, 1995).
4. Cystic Fibrosis Genetic Analysis Consortium *Hum. Mutat.* **4**, 167–177 (1994).
5. Forrest, S. *et al.* *Nature Genet.* **10**, 375–376 (1995).

* To whom correspondence should be addressed.

Bomb signals in old Antarctic brachiopods

SIR — Skeletal check-marks are commonly used to assess the age and growth of organisms. They are usually assumed to be formed annually. By using radiocarbon bomb signals to calibrate growth checks in shells of Antarctic brachiopods, we show that they were laid down with a sub-biennial periodicity. The data also indicate that low Southern Ocean ¹⁴C signals are probably not caused by upwelling deep water, but are more probably due to reduced atmospheric supply and long-term radiocarbon deposition in ice.

Articulate brachiopods dominate the macrofossil record from Cambrian times. They are sessile filter-feeders, present in all oceans of the world, and are classified as 'low-energy lifestyle' organisms¹. Growth-rate investigations, which are performed rarely, are typically based on analyses of bands in their calcium carbonate shell valves². All species with carbonate skeletons incorporate ¹⁴C from the environment. Radiocarbon concentrations in oceans are low compared with atmospheric and terrestrial levels, because their dissolved carbon pool is around 60 times the atmospheric pool. Variability in oceanic signals has been used to indicate areas of upwelling water³, and coral and bivalve mollusc skeletal signals to assess ¹⁴C records in latitudes to 42 °N (ref. 4).

SYNTHETIC DNA FRAGMENTS OF THE CYSTIC FIBROSIS GENE

DNA fragment	Sequence*	Mutation
WT	5'-{a}-A TAT CAT CTT TGG TG-{b}-3'	None
Δ F508	5'-{a}-A TAT CAT ... TGG TG-{b}-3'	Deletion of three bases
MU1	5'-{a}-A TAT <u>CGT</u> CTT TGG TG-{b}-3'	One-base substitution (underlined)
MU2	5'-{a}-A TAT <u>ACT</u> CTT TGG TG-{b}-3'	Two-base substitution (underlined)

The PNA used has the sequence H-CA CCA AAG ATG ATA T-Lys=NH₂.

*{a}, TGG CAC CAT TAA AGA AA; {b}, T TTC CTA TGA TGA ATA TA.

Antarctic ^{14}C levels are especially low, posing problems when dating marine material. Age corrections up to 1,400 years are necessary for recent material, whereas around 400 years is typical elsewhere⁵.

We obtained age estimates for the brachiopod *Liothyrella uva* from Signy Island, Antarctica (60°43' S, 45°36' W), by counting alpha growth checks in shell valves (G_c) and from measured increments (G_i) in a 2-year mark-recapture growth study. The G_i ages were 1.84 times older than G_c values (a in the figure), and this factor was significantly different from 2 ($t=15.8$, $P<0.001$). Ten brachiopods were then aged using $G_c \times 1.84$, and samples were cut from the oldest parts of their shell valves. Shell samples thus obtained were deposited between 1945 and 1985, and contained enhanced ^{14}C from the mid-1950s onwards (b in the figure). The ^{14}C increase occurred within 1–2 years of the first radiocarbon bomb signals found in nearby Antarctic peninsula ice-cores⁶. The ^{14}C signals peaked around 1962 and declined thereafter, although another peak possibly occurred between 1975 and 1980, following enhanced ice-core activity in 1972–74 (ref. 6).

Shell-activity data are, therefore, 2–5 years later than terrestrial signals, peaks are less precise, and the time required for the signal to return to background levels is much longer. This reflects the large carbon pool and long residence time in Antarctic circumpolar water, and contradicts the hypothesis that upwelling deep water with low radiocarbon content causes reduced signals in polar oceans³, which would cause signals to decay more rapidly (upwelling is

also unlikely from oceanographic considerations—Signy lies within the strongly stratified Weddell–Scotia confluence⁷). Lower marine signals in polar latitudes are, therefore, more probably due to reduced atmospheric ^{14}C supply and long-term radioisotope incorporation in ice.

The close correlation of shell age with ^{14}C signal confirms that growth checks were not formed annually, and indicates either entrainment to a biennial environmental cycle that occasionally breaks down, or an endogenous cycle with a periodicity of less than 2 years. Using the $\times 1.84$ conversion of growth rings to age gives a maximum of 52 years for *L. uva* (2 years' periodicity gives 56 years). Temperate and tropical brachiopod ages are generally 2–15 years. Ages >50 years were calculated recently for *Magellania fragilis*², a Weddell Sea species, assuming annual growth checks. Periodicities approximating 2 years would predict ages of 90–100 years.

Sea-ice dynamics in the Weddell Sea have a 7–9-year cycle⁸. Poor phytoplankton years also occurred at Signy Island on four occasions between 1972 and 1982 (ref. 9). However, these were not at approximately 2-yearly intervals, and no known sea-ice cycle approximates 2 years. Other long-term cycles identified include 4–6-year periodicities in life-history traits of phocids around the Antarctic Peninsula, which were suggested as being related to the El Niño–Southern Oscillation. However, the trends in these data were weak, and different biological cycles occurred in different phases with the El Niño signal¹⁰. A clear environmental signal of shorter periodicity has yet to be identified in this area. Alter-

natively, approximately biennial growth checks could result from endogenous cycles. Several Antarctic benthic invertebrates have reproductive cycles of >1 year (ref. 11), and reduced overall resource supply may underpin such extended cycles. Clearly, care is needed in interpreting shell check-marks and ageing cold water species, without independent dating. Whether the underlying causes of near-biennial growth bands are taxon-specific or environmentally modulated also requires elucidation.

Lloyd S. Peck

Natural Environment Research Council,
British Antarctic Survey, High Cross,
Madingley Road, Cambridge CB3 0ET, UK
Thomas Brey

Alfred Wegener Institute for Polar and
Marine Research,
D-2850 Bremerhaven, Germany

1. Peck, L. S. et al. *Treatise on Invertebrate Palaeontology (H)*, Brachiopods Ch. III (Geol. Soc. Am. & Univ. Kansas, Lawrence, in the press).
2. Brey, T. et al. *J. mar. Biol. Ass. U.K.* **75**, 857–869 (1995).
3. Broecker, W. S. & Peng, T. S. *Nature* **356**, 587–589 (1992).
4. Weidman, C. R. & Jones, G. A. *J. geophys. Res.* **98**, 14577–14588 (1993).
5. Gordon, J. E. & Harkness, D. D. *Quat. Sci. Rev.* **11**, 697–708 (1992).
6. Peel, D. A. & Clausen, H. B. *J. Glaciol.* **28**, 43–55 (1982).
7. Muench, R. D., Gunn, J. T. & Husby, D. M. *J. geophys. Res.* **95**, 18177–18190 (1990).
8. Murphy, E. J. et al. *Deep Sea Res.* **42**, 1045–1062 (1995).
9. Clarke, A., Holmes, L. J. & White, M. G. *Br. Antarct. Surv. Bull.* **80**, 65–86 (1988).
10. Ward Testa, J. et al. *Can. J. Fish. Aquat. Sci.* **48**, 631–639 (1991).
11. Clarke, A. *Invert. Reprod. Dev.* **22**, 176–184 (1992).

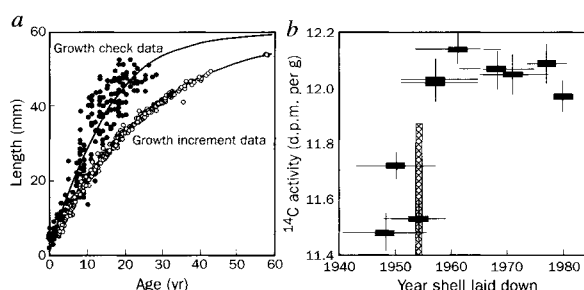
Mating success of male birds

SIR — Widemo and Owens¹ examined the amount of variation in mating success on mating arenas (leks) among male birds of the ruff *Philomachus pugnax*, and concluded that the skew in male mating success decreases with increasing average lek size. Such a relationship has important implications for our understanding of the evolution of lekking as a mating system and sexual selection in general^{1–3}.

Widemo and Owens' conclusion was challenged by Mackenzie et al.² on the basis that the skew index^{4,5} used to quantify the amount of variation in mating success between males is expected to decrease with lek size under a null version of the model in which mating occurs entirely at random. Although it is true that the skew index decreases with increased lek size under random mating (dashed line, a in the figure), this decrease is much less pronounced than the exponential decrease suggested by Mackenzie et al.² (solid line, a in the figure). The drop in skew observed by Widemo and Owens in larger leks is therefore not simply an artefact of the relationship between the skew index and lek size.

Two other possible factors could

a , Brachiopod shell length compared with age from external shell check-marks (G_c) and from growth increments (G_i) from a field mark-recapture study. Shell check-marks were analysed twice in 113 individuals by both authors independently (size range, 3–55 mm length; total, $n=205$).



For G_i data, 286 *L. uva* (size range: 4–54 mm length) were collected, measured and returned to the collection site. They were remeasured after 2 yr. General von Bertalanffy functions were fitted to both data sets (G_c data: $K=0.086$, $D=1.582$, $t_0=-1.834$, $R^2=0.873$; G_i data: $K=0.040$, $D=0.991$, $R^2=0.996$; L_∞ was set to 60.0 mm to cover the whole size range) using the SIMPLEX iterative nonlinear fitting algorithm. The difference between the two functions fitted a linear model (age (G_i) = $-0.785 + 1.842 \times \text{age} (G_c)$; $r^2=0.997$; $n=113$). After removing handling effects, by comparison with 6-month-duration trials (data not shown) at the same locality, mortality in the G_i experiment was $<2\% \text{ yr}^{-1}$. Growth rates in shorter-duration trials were also not significantly different from the 2-yr experiment. b , Shell ^{14}C activity (decompositions $\text{min}^{-1} \text{ g}^{-1}$, $\pm$ s.d.) versus age. The oldest 5–8 mm (corresponding to approximately the first 3–4 yr of life, calculated from G_i data in a) of the brachial valves of 10 *L. uva* 12–57 mm long were analysed by accelerated mass spectroscopy. They had 8–25 growth rings (15–46 yr old, using the $\times 1.84$ factor). Horizontal bars, 3–4-yr period of shell deposition; horizontal lines, 95% confidence interval for sample ages; hatched vertical bar, earliest enhanced radioisotope signals in Antarctic Peninsula ice-cores. ^{14}C data for samples before the ice-core signal were significantly lower than later ones (ANOVA, $F=76.5$; 2, 6 d.f.; $P<0.001$). Shell-activity levels correspond to ^{14}C ages of 890–1,340 yr.

Reconstruction of museum specimens

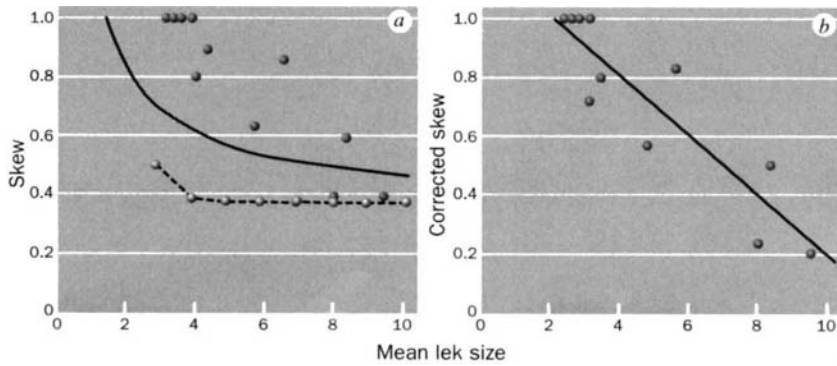
SIR — Museum collections of fishes, amphibians and reptiles have traditionally been repositories for single, fluid-fixed specimens (holotypes) on which most type-descriptions are based. These specimens are generally unsuitable for molecular studies because they have been fixed in formalin. Their value could be greatly enhanced by the use of non-invasive imaging techniques to provide structural information for the study of taxonomy, myology, functional morphology and also new characters for phylogenetic studies.

With the advent of high-resolution magnetic resonance imaging (MRI) scanners and reduction in the cost of scanning (£150 per specimen for the studies reported here), it is now feasible to use these methods for soft-tissue studies of smaller zoological specimens. As examples, we have used preserved elasmobranch fishes imaged using a high-resolution (4.7-T) scanner with a 20-mT m⁻¹ gradient set. Our comparisons between preserved and fresh material show that formalin fixation itself has no effect on image characteristics.

Standard, non-invasive techniques (radiography, whole-body staining) can give only limited anatomical information, because of the low radiographic contrast of the wholly cartilaginous, loosely articulated skeleton and the logistical difficulties of staining these large fishes, especially such multi-articulated structures as the branchial arches. In addition, neither technique can reveal details of myology, which can be studied only by dissection. Elasmobranch hyaline cartilage is friable and easily deformed during maceration, as osteocytes are absent and structural apatite mineralization is superficial except in the vertebral centra of some orders¹.

Counts of vertebrae are important at the specific level of elasmobranch taxonomy. There have been difficulties in the X-ray determination of vertebral counts². 'Notochordal sharks' are deep-sea sharks with reduced calcification, which prevents accurate radiographic determination of complete counts. *In situ* enumeration of vertebral centra was investigated with two-dimensional MRI imaging. A representative section of the anterior vertebral column of the bramble shark *Echinorhinus brucus* is shown in Fig. 1a overleaf. Segmentation of the notochord occurs internally, although there is little external evidence for it in the notochordal sheath. This is the first presentation of external and internal structure of the spinal column of *E. brucus*, and shows that vertebral counts using MRI are now feasible in these elasmobranchs.

Using the ANALYZE software³, three-dimensional reconstruction from two-



a, Observed relationship across leks between mean lek size and skew in male success. The mating skew index is calculated as the ratio of observed to maximum skew^{3,4}, which is becoming a commonly used measure of heterogeneity in reproductive success among group members (for example, refs 6–8). The dashed line is the expected relationship when females select males randomly (the null model). The expected distribution assumes that the average mating success of males is unity. Solid line, expected relationship proposed in ref. 2 under the hypothesis of random mating. The discrepancy is due to an error in ref. 2, partly caused by a problem with the random number generator (A. Mackenzie, personal communication) and most importantly by the inclusion of lek sizes smaller than two while estimating the effect of lek size variation on skew. Skew should not be estimated for leks smaller than two, and indeed, at least two males were always present when females visited leks (F. Widemo, personal communication). **b**, Observed relationship between mean lek size and 'corrected' skew when controlling for the effect of average mating success and the interdependence of the skew index on the number of males ($r^2 = 0.83$; $n = 11$; $P < 0.0001$). Expected skew under random mating (random skew) was estimated for each of the 11 leks by considering the number of males and females visiting a lek. The corrected skew quantifies the difference between the observed skew and the expected skew under random mating: corrected skew = (observed skew – random skew) / (maximum skew – random skew), where maximum skew = 1. A computer program for calculating the expected skew under the hypothesis of random mating (for leks with variable number (> 2) of males and females) is available on request from the authors (e-mail: lkeller@ulys.unil.ch).

explain the relationship observed by Widemo and Owens under the null hypothesis of random mating. First, greater skew is expected to characterize smaller leks if these were visited more frequently by males spending only a little time on the lek. This suggestion predicts a negative association between lek size and the average proportion of time that males are at the lek site. However, the correlation between these two variables is not significant ($r^2 = 0.02$, $n = 11$, $P > 0.05$; data provided by Widemo and Owens) and the sign of the correlation is positive, indicating that variation in the time that males spend at leks of different sizes is not responsible for the observed relationship between mating skew and lek size.

Second, as Mackenzie *et al.*² correctly point out, the skew index is sensitive to the mean mating success. As with any other formula for calculating skewness, the skew index tends to decrease with increasing mean mating success. Therefore, if mean male mating success increases with increasing lek size, we expect a negative relationship between skew and lek size even if mating is random.

To test this hypothesis, we examined the relationship between mean lek size and mating skew, while controlling for the effects of variation in mean male mating success and the interdependence of the skew index on the number of males (data provided by Widemo and Owens). For each of the 11 leks, we estimated the expected skew under the null hypothesis

of random mating and quantified the deviation between this value and the observed skew (corrected skew; *b* in the figure). This analysis reveals a strong and significant negative relationship between the corrected skew and mean lek size.

In conclusion, Mackenzie *et al.*² are right to point out that the skew index can be affected by factors such as lek size and mean mating success, even under the null hypothesis of matings occurring entirely at random. However, our analysis demonstrates that these factors do not account for the relationship that Widemo and Owens describe, thus validating the conclusion that mating skew decreases with lek size in the ruff.

Laurent Keller*

Michael J. B. Krieger

*Institut de Zoologie et d'Ecologie Animale,
University of Lausanne, BB,
1015 Lausanne, Switzerland*

1. Widemo, F. & Owens, I. P. F. *Nature* **373**, 148–151 (1995).
2. Mackenzie, A. *et al.* *Nature* **336**, 471 (1995).
3. Höglund, J. & Alatalo, R. V. *Leks* (Princeton Univ. Press, 1995).
4. Keller, L. & Vargo, E. L. in *Queen Number and Sociality in Insects* (ed. Keller, L.) 16–44 (Oxford Univ. Press, 1993).
5. Reeve, H. K. & Ratnieks, F. L. W. in *Queen Number and Sociality in Insects* (ed. Keller, L.) 45–85 (Oxford Univ. Press, 1993).
6. Keller, L. & Reeve, H. K. *Trends Ecol. Evol.* **9**, 98–102 (1994).
7. Bourke, A. F. G. & Heinze, J. *Phil. Trans. R. Soc.* **B345**, 359–372 (1994).
8. Reeve, H. K. & Keller, L. *Am. Nat.* **145**, 119–132 (1995).

*Also at Zoologisches Institut, Bern University, Ethologische Station Hasli, 3032 Hinterkappelen, Switzerland.

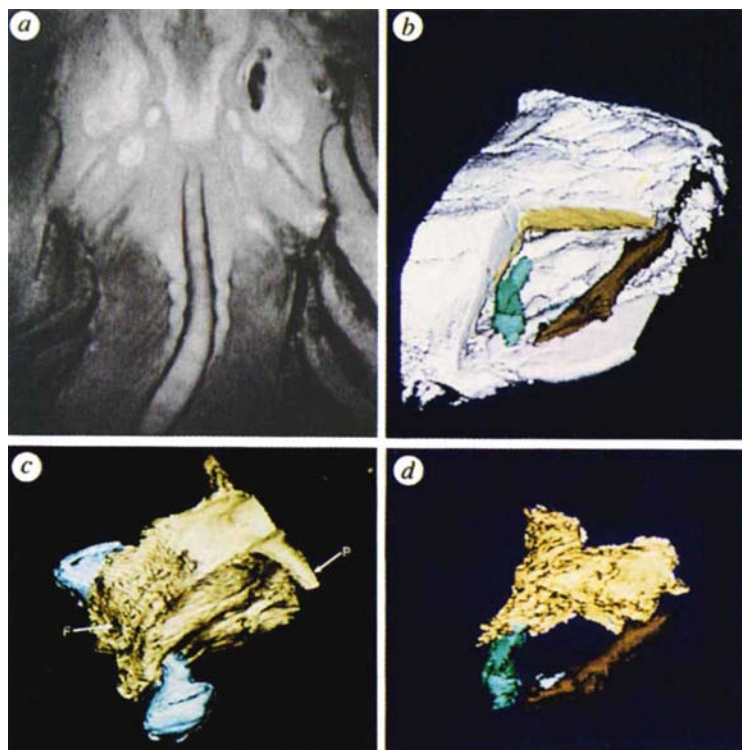


FIG. 1 a, Two-dimensional image, acquired as previously described⁶, of the head of *E. brucus*, showing notochord with internal segmentation. b, Image of computer-based three-dimensional reconstruction of the head of *O. maculatus*, showing part view with components selectively removed or shaded by computer. Spiracular cartilage (blue), palatoquadrate (brown), neurocranium (yellow), hyomandibula (green). c, Three-dimensional reconstruction of posterior half of neurocranium of *D. licha*, showing hyomandibulae (blue). P, postorbital process; F, foramen magnum. d, Three-dimensional reconstruction of cranial cartilages of *O. maculatus* with surrounding components selectively removed by computer and rotated to same orientation as b. Nasal capsules omitted.

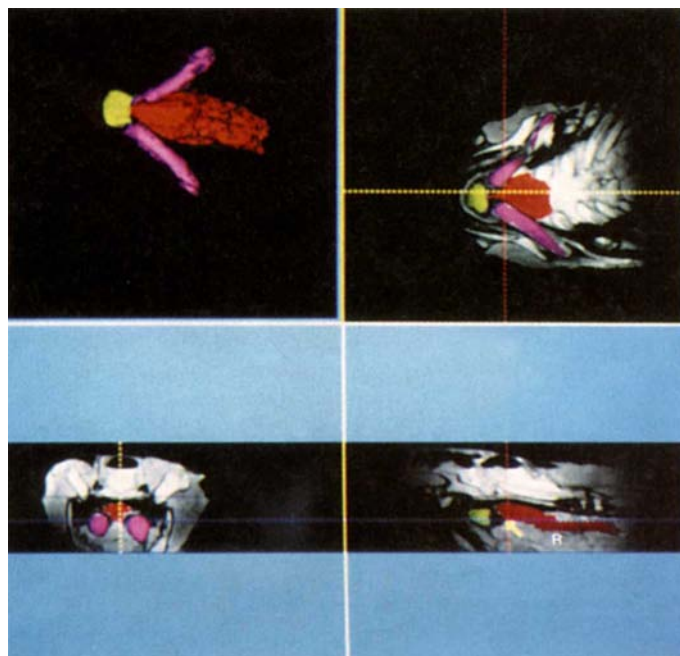


FIG. 2 Computer-based, three-dimensional reconstruction display of myology of *D. licha*, shown with associated orthogonal views. Basihyal retractor muscle (red) shown in top left in correct topographical orientation relative to ceratohyal cartilages (purple) and basihyal cartilage (green). Top right, horizontal section; bottom left, transverse section; bottom right, sagittal section (R, rectus cervicis muscle).

dimensional MRI images of the wobbegong shark *Orectolobus maculatus* showed the spiracular cartilage to be a single, plate-like structure lying in the anterior wall of the spiracle (Fig. 1b, d). Assessment of inter-specific variation revealed differences in the three-dimensional structure of spiracular cartilage; morphology can be visualized in three dimensions *in situ* without removing tissues from the specimen for separate dissection and staining. Our observations (to be published in detail elsewhere) suggest a previously unrecognized grouping uniting the Squalidae, Oxynotidae, Pristiophoridae and all batoids.

Simultaneous *in situ* analysis³ of shark muscle and cartilage is shown in a three-dimensional reconstruction of the neurocranium (Fig. 1c) and hyoid arch myology of the black shark *Dalatias licha*. The hypobranchial musculature, which lies in the region enclosed by and behind the lower jaw and is in close juxtaposition with the hyoid arch itself, is dominated by the elaborate rectus cervicis muscle complex. In addition to coracohyoid and arcualis muscles, a deep muscle bundle inserts on the posterior edge of the basihyal cartilage. Previously unrecognized or considered as a coracobranchial muscle⁴, the reconstructions presented here (Fig. 2) clearly show its topographical derivation from the rectus cervicis.

This new muscle (retractor basihyal muscle) shares a common origin at the pericardium with the rectus cervicis, but has a separate insertion on the basihyal (caudal edge as opposed to rostral edge). This arrangement helps jaw opening by widening the gape and lowering the floor of the oro-branchial cavity, which may enable *Dalatias* to feed on larger fishes; stomach contents contain whole small fishes but only chunks of larger fishes bitten out of the backs of the prey⁵.

The application of this technique to other vertebrate groups has great potential. Large stores of fixed material from numerous public and private collections are now candidates for nondestructive soft-tissue study, thus increasing their value as scientific resources, especially in cases where rare species are represented by a few extant specimens or when the holotype of a taxon is the only known specimen.

Geoffrey Waller*, John Cookson

Department of Morbid Anatomy,
The London Hospital Medical College,
Turner Street, London E1 2AD, UK

1. Ridewood, W. G. *Phil. Trans. R. Soc. B* **210**, 311–407 (1921).
2. Springer, V. G. & Garrick, J. A. F. *Proc. U.S. natn. Mus.* **116**, 73–96 (1964).
3. Robb, R. A. & Hanson, D. P. *Proc 1st Conf. Visualisation biomed. Computing* 507–515 (IEEE Computer Soc. Press, Atlanta, 1990).
4. Shirai, S. & Nakaya, K. *Zool. Sci.* **9**, 811–821 (1992).
5. Matallanas, J. *J. Fish Biol.* **20**, 155–163 (1982).
6. Waller, G. N. H. *et al. Magn. Res. Imaging* **12**, 535–539 (1994).

*Also at: Department of Zoology, The Natural History Museum, London SW7 5BD, UK.

Alone in the Universe?

Woodruff T. Sullivan III

Extraterrestrial Intelligence. By Jean Heidmann. Cambridge University Press: 1995. Pp. 235. £16.95, \$24.95.

Extraterrestrials: Where are They? Second Edition. Edited by Ben Zuckerman and Michael H. Hart. Cambridge University Press: 1995. Pp. 239. £25, \$39.95 (hbk); £11.95, \$19.95 (pbk).

It is indeed sobering that, after almost four decades of scientific investigations into extraterrestrial life, the field of exobiology (or bioastronomy) thrives despite its inability to demonstrate that its subject matter even exists. The modern era began in 1958 with the suggestion (in *Nature*) by Giuseppe Cocconi and Philip Morrison that a search should be made at radio wavelengths for signals originated by extraterrestrial intelligence. There have been more than a hundred searches (mostly at radio wavelengths), a few of which have been multi-year efforts sweeping the entire sky at specific frequencies. In 1976, a more direct approach was taken when the Viking spacecraft searched for evidence of organic chemistry in the Martian soil. Although nothing has yet been found, our Copernican sensibilities continue to foster the idea that we are unlikely to be special.

Extraterrestrial Intelligence by Jean Heidmann, an astronomer at the Paris Observatory, Meudon, is an excellent translation of the 1992 French original. Aimed at the general public, it succinctly covers all the important aspects of the field. In the first half, Heidmann takes the reader all the way from the Big Bang to the formation of the Solar System and the origin and evolution of life. He then deals in the second half more specifically with extraterrestrial intelligence — what it might be like and how one might search for it. Actual searches are described right up to Project Phoenix, the aptly named successor to the NASA project that the US Congress axed in 1993, as well as Projects Serendip and BETA, which now each simultaneously sample 200 million frequency channels. Heidmann writes in a lucid, agreeable personal style, but the lack of editing for an international audience seems peculiar — one is ever conscious that the book was originally written for readers in France (for example, Fourier transforms cannot be mentioned without referring to Baron Joseph Fourier himself). Also unusual for a popular book is the lack of illustrations, but this does not noticeably

detract from its accessibility.

Extraterrestrials: Where are They? edited by the astronomers Ben Zuckerman and Michael Hart stems from a 1978 symposium, but in this expanded second edition about a half of the text is either revised or

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Artist's impression of a remote-controlled robot for lunar mineral prospecting. In *Moon Handbook: A 21st Century Travel Guide*, Carl Koppeschaar weaves fact with fiction in an imaginative excursion around Earth's nearest neighbour in the year 2020. Moon Publications, \$10, £6.95 (pbk).

entirely new. Almost half of the individual chapters are moderately technical, the others more accessible for nonexperts. Authoritative, precise and well-written treatments are given for topics including the origin of life, evolution, interstellar travel and colonization techniques, UFO debunking, and search strategies for extraterrestrial intelligence. The tenor of the volume is sceptical about the idea that extraterrestrial intelligence exists at all, although more optimistic views are fairly

represented. The central argument against the existence of extraterrestrial intelligence, sometimes called the Fermi paradox, is that, if it is abundant in the Milky Way, then surely at least one technical civilization would have embarked on interstellar colonization and would now indeed pervade the entire Galaxy. Since "They" evidently are not now in our planetary system, then they must not exist at all.

The evolutionary biologist Ernst Mayr rails against physicists and astronomers who are optimistic about the existence of extraterrestrial intelligence, arguing that their biological naivety has misled them, and that the probability of intelligence and radio-bearing culture arising on an Earth-like planet is infinitesimally small. In fact, much of the intractability of estimating the likelihood of other life forms is precisely because we have of course only one example and have no robust way to estimate its *a priori* probability.

Pat Rawlings/NASA

The cosmologist Richard Gott intriguingly uses the Copernican principle (as in his article in *Nature* 363, 315; 1993) to estimate that the number of radio-transmitting civilizations in the Milky Way may be as high as 100 (with 95 per cent confidence), implying that our nearest neighbour is relatively distant and will not be detected unless its power levels greatly exceed ours.

The eleven chapters that remain unrevised from the first edition are in general not out of date, with the exception of Hart's on the habitability zone for planets around a star. In a 1979 study he argued that life-supporting conditions could last for billions of years on an Earth in only a razor-thin range of distances from the Sun (5 per cent of that of the Earth from the Sun), but recent work by J. Kasting and colleagues has widened this zone by a factor of roughly ten.

As we learn more and more about the Universe, the debates and research on the origin of life, extraterrestrial life and extraterrestrial intelligence will but grow. Discoveries such as that of the planets recently detected orbiting other Sun-like stars only whet our appetite and spur us on to scientific exploration of this fundamental question. In the sixteenth century, Europeans asked "Could it be that we are not at the centre of the Universe?". Four hundred years later, humans similarly ask "Could it be that we are not alone?" □

Woodruff T. Sullivan III is in the Department of Astronomy, University of Washington, Seattle, Washington 98195, USA.

Cosmic imagination

Giovanni F. Bignami

Imago Mundi: La rappresentazione del Cosmo attraverso i secoli. By Francesco Bertola. Biblos, Padua: 1995. Pp. 231. L80,000.

It is always a good idea to go back to the etymology of a word to appreciate its full meaning. 'Mundus' is Latin for 'clean' and 'beautiful', as well as referring of course to our world. And the Greek 'kosmos', the common root of 'cosmic' and 'cosmetic', also includes the idea of order: rather like saying, in one word, that the beauty of the

not to mention geography.

Bertola's *Imago* is a superb addition to this rich heritage. The author uses the continuum of cosmology, a constant of human interest, to view scientific culture as part of culture as a whole, including some of the finest examples of visual art. Take, for instance, the fresco by Giusto de' Menabuoi on the dome of the baptistry in the Scrovegni Chapel, Padua, not so far from Giotto's painting of comet Halley. Dating to 1300, it is a complete cosmo-logical representation of the Earth and its system of celestial spheres. And little more than a century later, at the beginning of the Renaissance and the Copernican 'revolutionary' era, Raphael paints in the Vatican the splendid "Allegory of Astronomy". Not only does he accurately depict the starry sky and its constellations, but he does so with a master's touch, endowing his heavenly picture with a sense of stereoscopic depth. Of the same period, and included in Bertola's book, is a paper woodcut by Albrecht Dürer of the constellations of the southern sky. Some of them are accurate and complete, but ample spaces remain at the extreme south (at the antipodes of Nuremberg). The woodcut was done, we learn, four years before Magellan's departure on his voyage of southern discovery. One and a half centuries later, astronomical knowledge had increased

ern view that stars are organized in vast systems, each of which contains an eye of God and is thus a "finite vision of infinity". The profound link between cosmos and mankind is apparent in the "Starry Self-Portrait" of Pistoletto (1973) which concludes the book. The artist's profile is etched on a true deep-sky star field: art and science live together in the human vision of cosmology. □

Giovanni F. Bignami is at the Istituto di Fisica Cosmica, 15 via Bassini, 20133 Milano, Italy, and Università di Cassino, Italy.

Eclipsed star

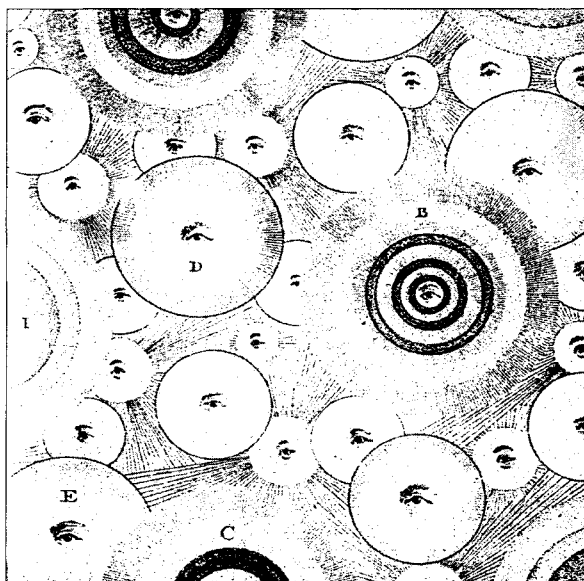
Owen Gingerich

The Correspondence of John Flamsteed, First Astronomer Royal. Edited by Eric G. Forbes, Lesley Murdin and Frances Willmoth. *Institute of Physics*: 1995. Pp. 955. £140, \$280.

IN 1835, Francis Baily, vice-president of the Royal Astronomical Society, shocked the British scientific establishment with the publication of *An Account of the Revd. John Flamsteed*. Perhaps the first scientific biography based on archival sources, Baily's account of Flamsteed, the first Astronomer Royal, showed Isaac Newton in a less than adulatory light. Newton's near-piracy of some of Flamsteed's astronomical observations revealed that the demigod of physics could play the part of the leviathan when it came to lesser mortals. "I am not fond of war", wrote the dour Astronomer Royal to a confidant.

Meanwhile, in the past few decades, Newton, who remains undisputed as the greatest natural philosopher and one of the three greatest mathematicians of all time, has been celebrated by a magnificent seven-volume edition of his correspondence and eight even thicker magisterial volumes of his mathematical papers, as well as several excellent biographies. Flamsteed remains, along with Halley, Hooke and Wren, a lesser light in that epoch-making era, but nevertheless an outstanding figure who founded the Greenwich Observatory and pioneered telescopic positional astronomy. Now, in the first of three thick volumes, Flamsteed's correspondence has been organized, translated when in Latin, and provided with a succinct commentary. The series is the work of the late Edinburgh historian Eric G. Forbes, and has been skilfully completed by the archivists Lesley Murdin and Frances Willmoth.

The years 1666 to 1682 take the young Flamsteed from his home in the North Country — Derbyshire — to London (where he temporarily resided in the Tower with his patron, Sir Jonas Moore)



Thomas Wright's star systems and the eye of God (1750).

Universe lies in its own order. The concept could be a fitting subtitle to Francesco Bertola's beautiful book dedicated to "imaging the cosmos over the centuries".

The title *Imago Mundi* itself has a fascinating history, worthy of a fiction by Umberto Eco. It first appeared in the Middle Ages in a codex by Onorius of Autun, somewhere between 1100 and 1150. Less than a century later, however, a treatise by Walter of Metz entitled *Image du Monde* became a big success, well before the invention of print. Hand-copied and richly illuminated, and rendered in both prose and verse (6,600 verses, all told), this work on astronomy and geography circulated in Hebrew, German and English as well as in the original mediaeval French. With the advent of printing, the book remained on the bestseller list for more than three centuries, despite fierce competition from the *Imago Mundi* (1483) of Pierre d'Ailly, Bishop of Cambrai, also a learned imager of astronomy and geography. Christopher Columbus is said to have been a keen reader of this book, but then he took astronomy very seriously,

greatly, as can be seen in the representation of the same portion of the sky by the Dutchman Andreas Cellarius. The book also contains an illustration from Kepler's *Astronomia Nova*, technical yet aesthetically pleasing, as well as, more recently, Escher's endlessly repeating symmetrical fishes (1959) and Seicker's violent picture of cosmic radiation (1990).

In his text, which is in both Italian and English, Bertola describes the history of ideas about the cosmos. By necessity this amounts to a list of contributions from individual thinkers, with each personality set in a historical context. The secret of easy and captivating reading is to let the individuals speak for themselves, and here they stand out vividly in their quotations. These range from Plato ("the true astronomer must be the wisest of all men") and Ptolemy through to Dante and the Copernican revolutionaries. The real revolutionary of that era, however, was Giordano Bruno, the first person to imagine that the "bodies which are beyond Saturn" are like the Sun, innumerable and surrounded by fainter planets. Or consider Thomas Wright (1750), with his mod-

and then to Greenwich and the building of the Greenwich Observatory (in 1675). Among the earliest letters preserved is a lengthy report on lunar occultations sent in 1669 by the 23-year-old Flamsteed to Lord Brouncker (president of the Royal Society). This brought him to the notice of the London establishment, and soon he was in active communication with Henry Oldenburg, secretary of the Royal Society and a redoubtable international correspondent. By 1673, Flamsteed was deemed to be the appropriate correspondent with J.-D. Cassini, who shortly before had become the effective director of the new Paris Observatory.

Flamsteed's star was clearly on the rise. He complained that his father would not buy a proper clock for him, and he asked Oldenburg if Riccioli's *Astronomia reformata* or Malvasia's *Ephemerides* could be purchased at a moderate price in London. By the spring of 1675, Flamsteed was writing — now from the Tower in London — to Halley, sending an analysis of Cassini's table of refractions as published in the Malvasia *Ephemerides*, so quite probably he got the books.

By 1676, now installed in Greenwich, Flamsteed writes to the most distinguished astronomical observer of the day, a man his senior by 35 years, Johannes Hevelius in Gdańsk. Gone is the highly deferential tone with which the young man had written to Cassini; in this letter, polite but assertive, he stands his ground on his criticism of Hevelius's naked-eye measurements. The correspondence is fundamentally defining for Flamsteed's research project, the measurement of stellar positions with telescopic sights and micrometers. Hevelius answers at length, sending a series of solstitial meridian observations of the Sun, claiming an accuracy better than 20 seconds of arc. Flamsteed proceeds to dissect these data, replying that they seem accurate at best to half a minute of arc.

Earlier, Robert Hooke had published his critical animadversions on Hevelius's methods, so the Gdańsk astronomer was on the defensive. This made Flamsteed and Hooke uncomfortable bedfellows, for Flamsteed had little respect for his abrasive colleague from the Royal Society. "Hee professes himselfe to be a Philosopher not an Astronomer and therefore I have no more to say to him but that astronomy is ill handled when it must be ordered by the Whimsies of Philosophy", wrote Flamsteed to Sir Jonas Moore on 2 April 1678. All this led to a trip by Edmond Halley to check out Hevelius's claims. Writing from "Dantzick" in 1679, Halley explained to Flamsteed that, "But

as the distances measured by the Sextans, I assure you that I was surpriz'd to see so near an agreement in them, and had I not seen, I could have scarce credited the Relation of any; Verily I have seen the same distance repeated severall times without any fallacy agree to 10' [sec-

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Precision observer: John Flamsteed.

onds]." Although the contest between Hevelius and Halley, who had brought along telescopic sights, proved to be a stand-off, Halley was never as skilled an observer as Flamsteed, whose methods pointed the way to the future of precision astronomy.

In 1680–81, this first volume of Flamsteed correspondence reaches its climax in an exchange with Newton. Still on friendly terms, Flamsteed proposes that the comets of November and December 1680 are one and the same, having gone into a hairpin curve in front of the Sun. Newton coolly suggests that he should defer the publication of his "hypothetical notions till they have been well considered both by his friends and himself". In fact, Newton allows that he suspects (erroneously) that the comets were indeed two different ones. The importance of this exchange is not explicitly mentioned in the present volume except to refer the reader to relevant modern discussions. Meanwhile, we are left in a state of high anticipation for the next volume in this important series. □

Owen Gingerich is at the Harvard-Smithsonian Center for Astrophysics, Cambridge, Massachusetts 02138, USA.

About time

John J. Tyson

Biochemical Oscillations and Cellular Rhythms: The Molecular Bases of Periodic and Chaotic Behaviour. By Albert Goldbeter. *Cambridge University Press: 1996. Pp. 605. £70, \$99.95.*

A BIOLOGICAL cell is a little bag of chemicals, but traditional notions of chemical reactivity are insufficient to understand how it works, because chemistry textbooks are dominated by the principles of equilibrium thermodynamics. The second law assures us that a closed chemical system at constant temperature and pressure will evolve eventually to a unique state of equilibrium that is temporally invariant and spatially uniform. A little bag of chemicals in that state would in actual fact be dead.

A living cell, on the contrary, is a dynamic entity continually changing in time and space, copying, signalling, reproducing, tearing down and rebuilding. These activities, so characteristic of life, take place in a strictly choreographed manner on the familiar stage of organelles, cytoskeleton, chromosomes and membranes. The rhythm of the dance is tapped out by certain ubiquitous oscillations: the reproductive cycle of DNA synthesis, mitosis and cell division, the once-a-second signalling of cardiac pacemaker cells, the periodic waves of calcium that sweep across egg cells, and the 24-hour rhythms that dominate our physiological responses throughout each day. Chemistry books are of little help in making sense of the dynamics of living cells.

Unfortunately, biochemistry books will not solve the problem either. Sequences, codes, folding, binding, recognition — the everyday parlance of molecular biologists is remarkably successful, but the brilliant light of the structure-function paradigm has blinded many practitioners to its fundamental weakness. Like equilibrium thermodynamics, it emphasizes stable structures and pays scant attention to the time dimension. The problem is not lack of interest in the temporal and spatial organization of cells, but lack of tools for approaching these problems systematically and reliably.

What are the proper tools? To describe how molecular and cellular systems change through time and space, we need to learn the basic rules of chemical reaction kinetics, molecular diffusion, convection, chemotaxis, mechanical stresses and strains, and so on. The rules are not mysterious, but they quickly immerse the serious student in a soup of complicated differential equations. To understand these equations (in other words, to tie molecular mechanisms to the behaviour

of intact cells) requires sophisticated mathematical reasoning and well-planned numerical simulations. Not many molecular biologists are inclined to such rigours, nor are many applied mathematicians conversant with the biological problems.

Albert Goldbeter is one of the few scientists in the world who speak both languages fluently. Here he communicates his style of analysing periodic behaviour of biochemical and cellular control systems. He treats in great detail all the classical cases — glycolytic oscillations, cyclic AMP signalling, pulsatile hormone secretion, calcium oscillations, cell cycles and circadian rhythms — and is a reliable guide through the intricacies of the biochemistry and mathematics. The great strength of the book is his unshrinking dedication to understanding these processes from start to finish. The volume is loaded with experimental data, hypothetical mechanisms, differential equations, numerical simulations and careful comparison between theory and observations. Although the book is conversational, clear and accurate, Goldbeter makes no concessions to his readers. He assumes equal familiarity with metabolic pathways, polyacrylamide gels, Hopf bifurcations and deterministic chaos. Few readers will be able to keep pace as he flips back and forth between biology and mathematics. This is a book to read with a friend who speaks the 'other' language.

My only complaint is that Goldbeter discusses in detail only his own models of each rhythmic process. I can forgive this in a man who practically invented the field, especially in someone as unpretentious and likeable as Goldbeter. But there are other models of some of these processes that are superior to his and should have been given top billing. On the other hand, Goldbeter is generous, thorough and tasteful in his citations to the literature, from which the reader can easily gain access to the best experimental and theoretical work available.

Although it will demand some hard thinking, the book can serve as a reliable and rewarding introduction to mathematical modelling of dynamic processes in living cells for any molecular biologist beginning to see the usefulness of this tool. As Michael Berridge says in his foreword to the book: "I [have begun] to appreciate the value of good modelling, especially now that we have learnt so much more about the basic biochemical details regulating cellular activity. For each system now being investigated, there are so many variables that it becomes impossible to use our intuition to assess how each parameter influences the oscillatory cycle. The only way to understand these biological rhythms, therefore, is to become more quantita-

tive and to develop rigorous mathematical models. Albert Goldbeter is at the very forefront of this new approach." □

John J. Tyson is in the Department of Biology, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0406, USA.

Shifting sands

Martin A. Conway

Memory Distortion: How Minds, Brains, and Societies Reconstruct the Past. Edited by Daniel L. Schacter. *Harvard University Press: 1996. Pp. 417. \$49.95, £31.50.*

A COMMONLY held belief is that long-term memory is like a library, a personal library that we each carry around in our heads. In remembering past events, we take books from the shelves and examine their contents. On other occasions we might misplace a volume. Later we might quite by accident stumble upon it and feel the sense of wonderment that accompanies the recall of a 'forgotten' event. It is an attractively simple and beguiling metaphor, but one that students of human memory have long known is wrong.

Nor is human memory like a photograph album, a collection of cassettes, compact discs or videos or any other accumulative archive of the past. Rather, memories are fragmentary, condensed, often distorted and inaccurate representations of past experience. This point is made in impressive detail by all the contributors to this excellent collection of essays on memory distortion. Memories carry with them personal meanings and are interpreted in terms of the current knowledge, understanding and goals of the rememberer or, in the case of collective memories, of the group or society. They take effort to construct and are difficult to hold consciously in mind for more than a few seconds: they quickly dissipate from awareness and must be laboriously brought back to mind. In short, they are transitory mental constructions that encompass knowledge at different levels of abstraction — from the specific to the general — drawn from long-term memory and from the current processing environment and held together by control processes in the frontal lobes of the brain.

Patients with neurological damage to the frontal lobes can no longer construct mental representations of the past. Instead, as Morris Moscovitch shows, these patients replace gaps left by their memory disorder with imaginary remembered experiences consistently believed to be true. The mental constructions that must mediate their memory reports are not shaped to give an

accurate account of past experience in even its most general aspects.

Long-term memory is also exquisitely sensitive to cues in the external environment and to knowledge internally generated or activated. We are, perhaps, constantly reminded of the past although we may not be directly aware of this. Control processes in the human brain modulate access to knowledge activated in long-term memory, allowing only some information to enter central processing sequences. Clearly, it would be no good if every cue one encountered led to the eventual construction of a memory (although Proust may have disagreed). On the other hand, being reminded of some relevant past experience can dramatically help in problem solving or achieving personal goals.

Larry Squire examines how the transitory but temporarily stable patterns of activation that make up memories may be distributed over anatomically and cognitively distinct memory systems that represent knowledge in different ways. James McClelland gives a preliminary outline of how memory construction might take place in parallel distributed processing architectures. Within this overall framework it is easy to see how memories can be accurate, inaccurate and incomplete all at the same time. The constructivist account also provides ready mechanisms for the incorporation of false information into memories and even the construction of wholly false memories, not only in patients with damage to their frontal lobes but also in normal people under certain experimental conditions.

In his scholarly and thorough introduction, Schacter traces the history of experimental memory distortion, and Elizabeth Loftus and colleagues and Stephen Ceci describe some recent exciting work on the creation of wholly false memories in normal people. Interestingly, only about a third of the participants in these experiments ever develop false memories, which indicates that there are individual and group differences in memory accuracy.

Arising from a seminar held by the Mind, Brain and Behavior Initiative at Harvard in May 1994, *Memory Distortion* provides an outstanding multidisciplinary perspective on memory accuracy, ranging from cognitive psychology through psychiatry, neuropsychology and neurobiology, to sociocultural analyses. My only reservation is that the book does not deal with the implications of a constructivist account of memory, particularly one emphasizing inaccuracy. If human memory is inaccurate and distorted, what use is it? My own view, for what it is worth, is that memory is primarily a vehicle for personal meanings and for grounding the self, and that accuracy is secondary to this role. □

Martin A. Conway is in the Department of Psychology, University of Bristol, Bristol BS8 1TN, UK.

State-dependent life histories

John M. McNamara & Alasdair I. Houston

Life-history theory is concerned with strategic decisions over an organism's lifetime. Evidence is accumulating about the way in which these decisions depend on the organism's physiological state and other components such as external circumstances. Phenotypic plasticity may be interpreted as an organism's response to its state. The quality of offspring may depend on the state and behaviour of the mother. Recent theoretical advances allow these and other state-dependent effects to be modelled within the same framework.

THERE is considerable diversity in the patterns of reproductive behaviour shown by different species^{1,2}. In some species, individuals do not reproduce until they have grown to a large size (for example, some trees), whereas in other species (for example, many insects and small mammals) there is rapid maturation. Some organisms reproduce once and then die (for example, Pacific salmon), others have several episodes of reproduction (most mammals). Life-history theory attempts to explain the scheduling of reproduction throughout an organism's lifespan, and hence to account for this diversity.

Life-history theory often assumes that fecundity and survival are functions of an organism's age. This framework has been used to analyse such questions as the age at which an organism should first reproduce and whether an organism should breed once or repeatedly^{1,2}. But if taken literally, the approach would amount to assuming that all individuals of the same age are equivalent, ignoring the underlying physiological and environmental conditions that determine survival and reproduction. We will refer to these conditions as the organism's state.

The limitations of the age-based approach to life histories have been known for some time³. Organisms generally differ in 'condition' or 'quality' and this is important in determining their reproductive success⁴⁻⁹. Work on the collared flycatcher¹⁰ is noteworthy in that manipulating a bird's clutch size can not only change the bird's future reproductive success, but can also influence the reproductive success of surviving offspring from the manipulated clutch, suggesting an underlying condition variable. Increasing a bird's clutch size results in the bird increasing its total parental effort, but not by enough to prevent a decrease in effort per offspring. The increase in total effort by the parent decreases its condition. This change in condition persists into future breeding seasons and manifests itself as a reduction in clutch size in these seasons. The decrease in effort per offspring results in offspring not only having lower survival but also being in poorer condition at maturity and hence having lower reproductive success. Thus even in a small animal, the effects of behaviour upon condition can last for several years, and a mother's behaviour can influence the condition of her offspring. But it is not yet clear what constitutes condition in the collared flycatcher. An interaction between the immune response to parasites and disease, nutritional status and reproductive effort may be responsible for the observed effects¹⁰.

Individual optimization of life-history decisions has been observed in other studies^{6,11,12}. For example, clutch size and lay date in kestrels depend in an adaptive way on the male's rate of catching prey⁶. The dangers of ignoring individual optimization are well illustrated by work on clutch size in birds. Within a population, the most productive clutch size may be larger than the modal clutch size². This has been regarded as a discrepancy that requires explanation¹³, but it is only a discrepancy if all breeding individuals of the same age are taken to be identical.

With individual differences, clutch size and reproductive success are likely to be correlated¹⁴.

A further limitation of the age-based approach is that it assumes that all surviving offspring are equivalent. As a result it cannot handle two important intergenerational effects: (1) the ability of a parent to increase the quality of its offspring by reducing the number of offspring produced (the number-quality trade-off), and (2) the existence of phenotypic correlations between parents and their offspring (for example, maternal effects). The number-quality trade-off has been analysed in a framework that looks at a single reproductive decision^{15,16} but this cannot be incorporated into an age-based approach to life histories. Models based on population genetics have been used to predict how maternal effects influence the change in phenotype from one generation to another¹⁷, but the adaptive consequences of maternal effects have been little explored and can only be understood within a state-based approach to life histories.

A state-based approach to life histories

An organism's ability to survive and reproduce may depend on its size and its condition, where condition may include, among other things, territory quality^{6,18}, fat reserves¹⁹, protein reserves²⁰, foraging skills²¹, parasite load^{22,23} and the state of its immune system¹⁰. For example, in red deer, male mating success depends on fighting ability, which is related to body size. The breeding success of females is dependent on body size, condition and the quality of the home range^{4,24}. For the purpose of modelling its life history, any of these or other features that are relevant should be included in the organism's state. It may also be convenient to include age as a component of state, although effects of age are really just consequences of underlying physiological state variables.

Although in some cases the effects of size can be included within the age-based approach²⁵, understanding the general effects of state requires a different framework. Population projection matrices provide one component of this framework. For a population structured by age alone, the Leslie matrix provides a means of projecting the size and composition of a population into the future. A state-based projection matrix gives the corresponding demography for a population structured by state²⁶. State-based matrices have been promoted by Caswell²⁷, and have been used for a variety of purposes, including analysis of community structure²⁸. They may be used to look at the growth rate of a whole population, but they can also be used to find the growth rate λ that results from following a particular life-history strategy. The growth rate then provides a measure of the fitness of the underlying strategy. The projection matrix for a particular species in given conditions does not in itself reveal anything about life-history evolution. To investigate evolutionary issues, it is necessary to determine how the growth rate λ changes under changes in the organism's life-history strategy. Caswell achieves this by investigating the sensitivity of λ to changes in the parameters of the projection matrix.

This approach has been used to determine the strength of selection acting on various life-history pathways in teasel²⁷ and killer whales²⁹. The sensitivity of λ to parameters also forms the basis of Lande's quantitative genetics justification of λ as a measure of fitness²⁷.

In this review we characterize state-dependent effects by the use of projection matrices, but in finding optimal life histories we will not use the sensitivity analysis of Caswell. Instead we favour a method that works directly with the number of descendants left far into the future. The mathematical basis for this method is provided by the theory of Markov decision processes. This theory deals with optimal state-dependent sequences of decisions, and has been successfully applied to the behaviour of organisms over part of their lives³⁰. The approach combines the theory of state-dependent projection matrices with the theory of Markov decision processes to provide a general state-based approach to life-history theory^{31,32}. Using the approach, it is simple to characterize and find optimal life histories^{31,33,34}. The approach represents life-history trade-offs through the effect of behaviour on future state.

A mechanistic basis for trade-offs. An organism can typically only increase its current reproduction at the cost of some loss in future reproduction. For example, this may be because its probability of survival is reduced or because the energy devoted to reproduction is not available for growth which would contribute to future reproduction. Trade-offs such as these are the basis for adaptive life-history decisions. In the state-based framework, trade-offs are mediated by an organism's state.

For simplicity we consider an organism that makes annual decisions about its reproductive behaviour during the coming year. A model is given by specifying four functions of the organism's state and behaviour which determine the following four quantities:

*f*1, The probability that the organism survives until next year.

*f*2, The state (or distribution of states) of the organism in one year's time, given that it survives until this time.

*f*3, The expected number of offspring produced which are alive in one year's time.

*f*4, The distribution of states of the offspring.

Each of the 45 trade-offs listed by Stearns² can be represented by an appropriate choice of these functions. For example, the trade-off between parental survival and current production of offspring can be represented by suitable choice of functions *f*1 and *f*3. The trade-off between offspring quality and number can be modelled by an appropriate choice of functions *f*3 and *f*4. Maternal effects can be modelled by allowing function *f*4 to depend on the state of the mother.

A life-history strategy specifies how the organism's reproductive behaviour depends on its state. If two individuals follow the same strategy, they need not be observed to perform the same behaviour, because behaviour depends on state and this may differ between the organisms. To assess a given strategy, we assume that an organism and all its descendants adopt this strategy, and look at the expected number of descendants left far into the future by the organism. The fitness of a strategy is measured by the asymptotic growth rate in descendant numbers³⁵. An optimal strategy maximizes this fitness measure and can be found by dynamic programming^{31,33,34}.

Having found the optimal strategy, it is possible to determine the relative advantage of being in one state as opposed to another. Take some reference state x_0 and define the reproductive value $V(x)$ of an individual in state x as the expected number of descendants left far into the future by the individual relative to the expected number left by an individual in state x_0 . As in the age-based theory³⁶, reproductive value can be used both to characterize the optimal strategy³⁷ and to give insights into adaptive reproductive decisions.

Models developed within this framework have looked at various

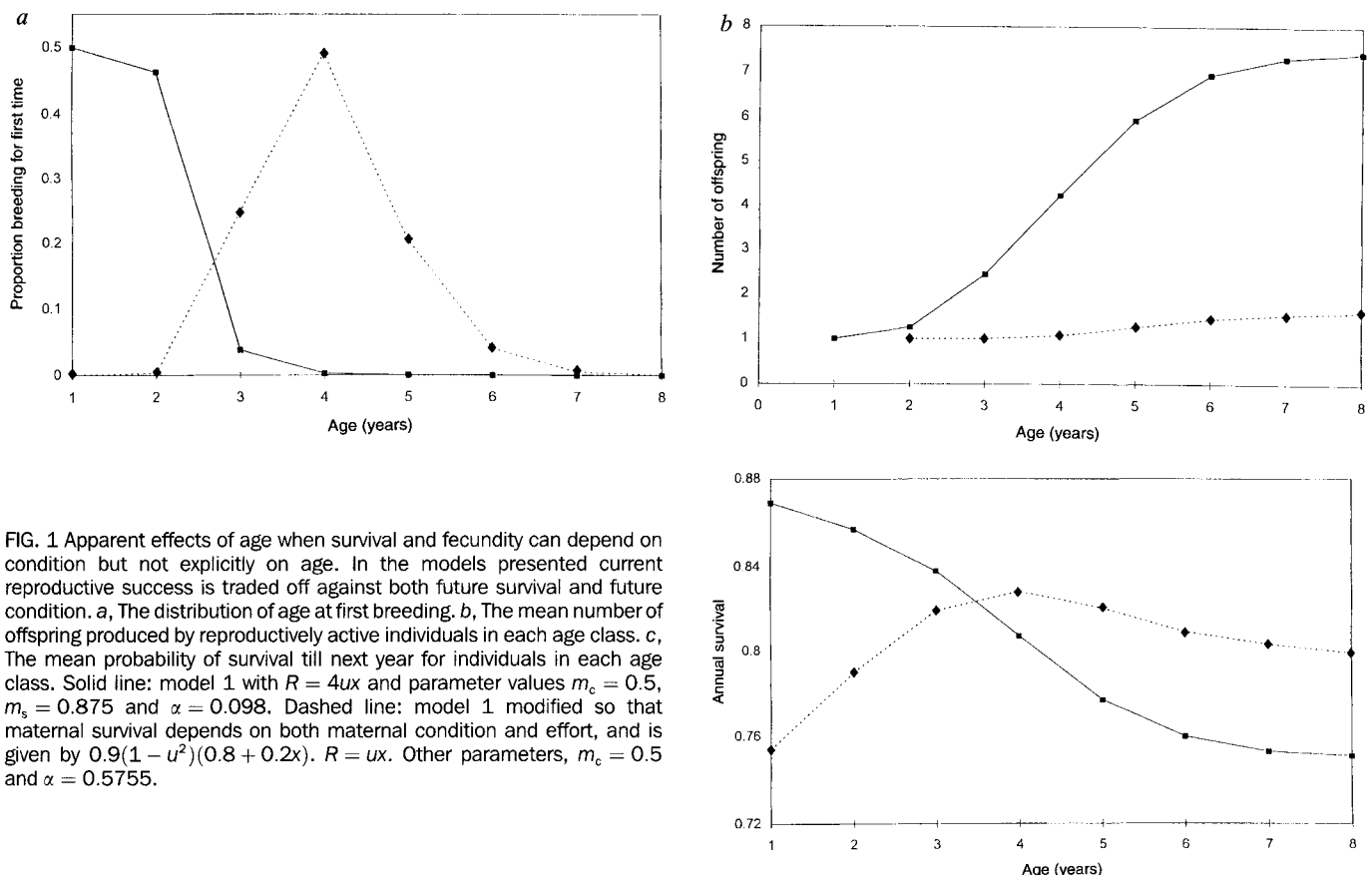


FIG. 1 Apparent effects of age when survival and fecundity can depend on condition but not explicitly on age. In the models presented current reproductive success is traded off against both future survival and future condition. *a*, The distribution of age at first breeding. *b*, The mean number of offspring produced by reproductively active individuals in each age class. *c*, The mean probability of survival till next year for individuals in each age class. Solid line: model 1 with $R = 4ux$ and parameter values $m_c = 0.5$, $m_s = 0.875$ and $\alpha = 0.098$. Dashed line: model 1 modified so that maternal survival depends on both maternal condition and effort, and is given by $0.9(1 - u^2)(0.8 + 0.2x)$. $R = ux$. Other parameters, $m_c = 0.5$ and $\alpha = 0.5755$.

topics. These include the offspring quality and number trade-off and maternal effects^{33,34}, phenotypic plasticity³⁸ and reproductive decisions in Soay sheep³⁹.

Apparent effects of age

A fundamental issue in life-history theory is the amount of effort that an organism should allocate to reproduction^{1,2}. The age-based approach to life-history theory looks for the optimal age at which to begin reproducing and the subsequent pattern of reproductive effort^{1,2}. The state-based approach looks for the best level of reproductive behaviour as a function of the organism's state. In general, state may include age as one of its components. Here we present a model (model 1) in which age has no effect *per se* and condition is the only state variable. Nevertheless, there are apparent effects of age because age and condition are correlated. **Model 1.** At the start of the breeding season animals differ in some measure of condition. During the breeding season a female makes two decisions. It first chooses whether to reproduce or not, and if it reproduces, it decides on the number of offspring to produce. The choice of offspring number can depend on the animal's condition. Having produced its young, the animal must choose the effort it puts into provisioning the young. This effort can depend on the animal's condition and the number of young for which it is caring. Increasing effort decreases maternal survival and future maternal condition, but increases the resources available to the young and hence the number of young which survive. The functions f_1 to f_4 which determine the trade-off are given in Table 1. None of the functions f_1 to f_4 depend explicitly on age. In this model, all offspring are in the same state if they survive to age 1 (and the population size is stable because of density-dependent effects: see Table 1) and maximization of the expected number of descendants left far into the future is equivalent to maximization of the expected number of surviving offspring that are produced over a lifetime.

Age of first reproduction. The optimal strategy involves not breeding when condition is below a certain critical level, and breeding when condition equals or exceeds this level. When the animal does not breed, its condition increases. When the animal's state at age 1 is below the critical level, there will therefore be delayed reproduction until condition rises sufficiently. Because there is a stochastic component to the annual change in condition, the delay will not necessarily be equal for all animals (Fig. 1a). Variation in the age of first reproduction occurs because of variation in condition at a given age⁴⁰. The dependence of the age of first reproduction on condition has been described in the moose⁴¹, the bison⁴² and the elephant seal⁴³; related effects have been found in red deer⁴⁴.

When there are effects of condition it may be misleading to look for a single age of first reproduction. Work on the elephant seal⁴³ illustrates this point. Females in the study population differ in their age of first breeding, in particular some breed at age 3 and some at age 4. In an attempt to understand this variation, Reiter and LeBoeuf⁴³ compared the fitness of two age-based life histories; one in which females always start breeding at 3 years, and the other

in which females always start breeding at 4 years. They conclude that at their study site, the optimal life-history strategy is to start breeding at age 4. They are consequently unsure why any female in the population should breed at age 3, and give a variety of possible explanations. There is, however, variation in the mass of females at weaning and heavier females tend to breed earlier than lighter females. This indicates that female life-history strategies cannot be understood by a purely age-based approach; some aspect of female condition must also be included.

Fecundity and effort over the lifetime. For the parameters used in Fig. 1, condition tends to increase with age. Because offspring number increases with condition, mean offspring number increases with age (Fig. 1b).

There has been much discussion of the relationship between reproductive effort and age^{1,2,45,46}. The standard argument is that effort should increase with age when reproductive value decreases with age. A decrease in reproductive value with age could occur when there is reduced fecundity or increased mortality as age increases. In our model both effort and reproductive value increase with condition. Thus in contrast to the standard argument, effort increases with reproductive value. Because mean condition increases with age, both mean effort and mean reproductive value increase with age.

Senescence. 'Ageing' or 'senescence' is used to refer to a drop in survival or reproduction later in an individual's life⁴⁷. In optimality theories of ageing, senescence is not necessarily inevitable and is the outcome of an optimal trade-off between current and future reproduction. It is usually envisaged that the trade-off is mediated by a conditional variable, and these theories can be modelled within the state-based framework. Indeed, Fig. 1c shows that model 1 predicts an increase in mortality with age. This occurs because reproductive effort increases with condition and hence with age. Similar effects occur in models based on size⁴⁸. Thus apparent senescence can occur without any deterioration in condition with age. In model 1, condition can increase or decrease. In most organisms there may be components of condition that deteriorate over time or at best remain the same.

TABLE 1 Specification of models 1–3

	State:	Models 1, 2 Model 3	condition, x ($0 \leq x \leq 1$), social status, x ($0 \leq x \leq 1$).
	Actions:	Model 1 Model 2 Model 3	(1) number of offspring to produce, n ($n = 0$, or $n \geq 1$). (2) effort put into provisioning the young, u ($0 \leq u \leq 1$). number of offspring to produce, n ($n = 0$, or $n \geq 1$). effort put into maintaining social status, u ($0 \leq u \leq 1$).
f_1 , Maternal survival		Model 1 Model 2 Model 3	$m_s(1 - u^2)$ does not survive $0.95(1 - u^2)$
f_2 , Maternal condition		Model 1 Model 2 Model 3	$x + 0.2 - m_c u^2$ not applicable $x + 0.25u - 0.125$
f_3 , Offspring survival		Models 1,2 Model 3	Let R be the total resource provided by the parent and let $r = R/n$ be the resource per young Let $S(r) = \alpha r^2 / (r^2 + 0.2^2)$. Then the expected number of offspring surviving is $nS(r)$. One offspring produced, survives with probability αx
f_4 , Offspring condition		Model 1 Model 2 Model 3	fixed at 0.1875. $0.5 + \text{slope}(x - 0.5) + \beta[(r^2 / (r^2 + 0.4^2)) - 0.5]$. maternal status, x .

In all computations, the parameter α has been adjusted so that the growth rate under the optimal strategy is 1. This is equivalent to assuming density dependence acting on the juveniles. Computations are carried out using a grid of 9 values of $x = 0, 0.125, 0.25, \dots, 1$ and using linear interpolation. The functions f_2 and f_4 give mean condition next year. Actual condition next year is stochastic and is found by first linearly interpolating the mean condition between adjacent grid points and then adding a small amount of additional stochasticity.

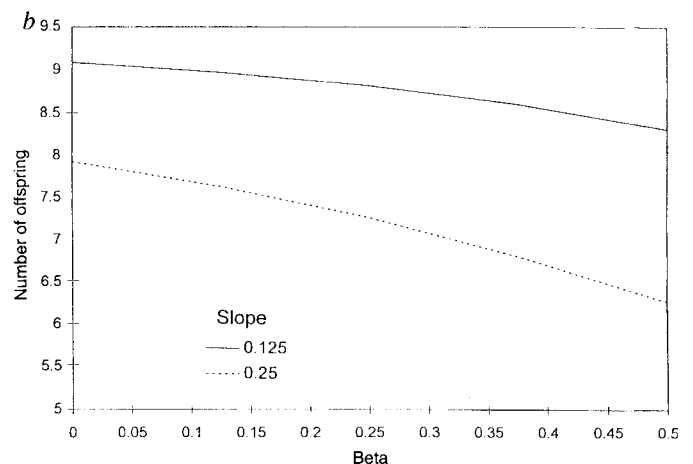
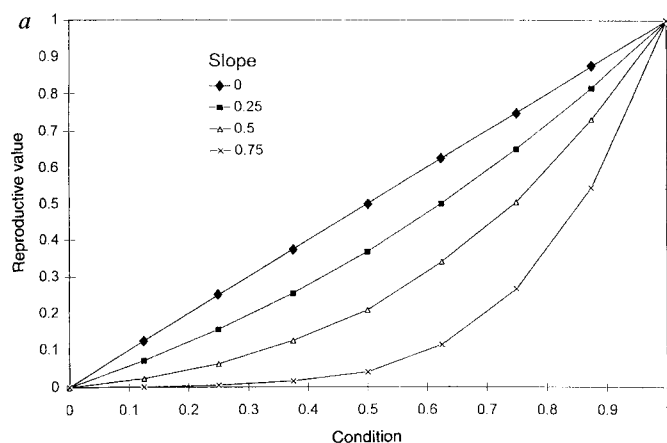


FIG. 2 Results from model 2 in which there are maternal effects and a trade-off between offspring number and quality. *a*, Reproductive value as a function of condition for various strengths of maternal effect. Reproductive value is measured relative to an individual in condition = 1. As slope increases, maternal effects become stronger (see text). $\beta = 0$. *b*, The

optimal number of offspring produced by an animal in condition 0.5. This number decreases as either the strength of maternal effects (slope) increases, or as the degree to which the animal can increase the condition of offspring by extra provisioning, β , increases. $R = 4ux$ throughout.

Realistic models of ageing need to include such variables. One approach (the disposable soma model) assumes a condition variable whose value is decreased by reproductive effort⁴⁹. If there is a component of mortality that does not depend on reproduction, then its magnitude will influence the optimal life history. High mortality favours early reproduction at the cost of a decrease in condition and hence results in early senescence^{47,49}.

Skipping reproduction. When effort may strongly reduce maternal condition in the following year, condition in the year after breeding may be below the critical level for breeding, and the animal will not reproduce. There will then be a negative correlation between whether an animal has bred in the previous year and whether it breeds in the current year. Such phenomena make no biological sense in models based solely on age. A comparison of the breeding pattern of the black-browed and the grey-headed albatross is instructive in this context. These species are very similar in body mass but differ in their pattern of reproduction. The black-browed albatross is an annual breeder. The grey-headed albatross does not breed in the year following a successful breeding attempt, but may breed after a failed attempt⁵⁰. Prince *et al.*⁵⁰ suggest that the longer breeding schedule and the poorer diet of the grey-headed albatross make it difficult to regain breeding condition after a successful breeding attempt.

Throughout this section, we have considered no explicit effects of age. We are of course not denying that survival or reproduction may be closely correlated with age. Although there is likely to be an underlying condition variable in such cases, it may be so strongly linked with age that we can work with age alone as a state variable. If, however, there are major differences between individuals of the same age, age cannot be the only state variable.

Intergenerational effects

When either (1) the mother's condition can change during her lifetime and there are maternal effects, or (2) there is a trade-off between the number of offspring produced and their quality, the state of offspring produced will depend on the mother's strategy. Under these circumstances, the fitness of a strategy cannot be assessed simply by counting the number of offspring produced over the mother's lifetime. Instead it is necessary to use a measure that takes the reproductive value of offspring into account.

Maternal effects are widespread, and grandmaternal effects have been found. If female rats are given a restricted diet, there is an impaired antibody response in first- and second-generation offspring⁵¹. Female rats were fed a low-protein diet for one month before mating and during pregnancy. Offspring were fed a normal

diet and mated to normal males. It was found that the resulting offspring had lower cerebral weight and lower cerebral DNA at birth⁵². In a field study of white-tailed deer⁵³, there was evidence of an effect of maternal and grandmaternal condition on mass and vulnerability to wolf predation. Second-generation effects have also been found in humans⁵⁴. Not all maternal effects will be based on physiological variables. Under some circumstances red squirrel mothers bequeath their territory to their offspring⁵⁵; here territory quality is a component of state that can be passed on to offspring. Maternal effects are well documented in insects⁵⁶ and plants⁵⁷. The trade-off between the number and quality of offspring is also widespread^{1,2}.

Model 2. This is a modification of model 1 which incorporates both maternal effects and the number-quality trade-off (details in Table 1). To concentrate on these effects, we assume that after reproduction the mother dies, so that generations do not overlap. Under this assumption it is optimal for the female to put the maximum effort into provisioning the young, and the female's only decision is the number of offspring to produce as a function of her condition. The effect of maternal allocation of resources and the effect of maternal condition act additively to determine the condition of the offspring at maturity (Table 1). The parameter β is the maximum amount by which the mother can increase the condition of an offspring by increasing the provisioning to it. The parameter slope is the slope of the regression of offspring condition on maternal condition when $\beta = 0$.

In this model it is assumed that the total resources allocated by the mother are proportional to her condition. It follows that in the absence of maternal effects (slope = 0) the reproductive value of an individual is proportional to its condition (Fig. 2a). When slope $\neq 0$, the reproductive value depends on the parameters slope and β , together with the reproductive behaviour. Figure 2a shows the effect of slope when $\beta = 0$ and behaviour is optimal. The figure shows that as the strength of maternal effects increase (increasing slope), reproductive value departs more and more from linearity, with the result that the value of high condition relative to low condition increases.

Because the parent can increase offspring condition by decreasing the number of offspring produced, the optimal number of offspring produced does not maximize the number of surviving offspring. Figure 2b illustrates the dependence of the optimal number of offspring on the parameters slope and β . Increasing either parameter for a given value of the other parameter increases the allocation per offspring and hence decreases offspring number.

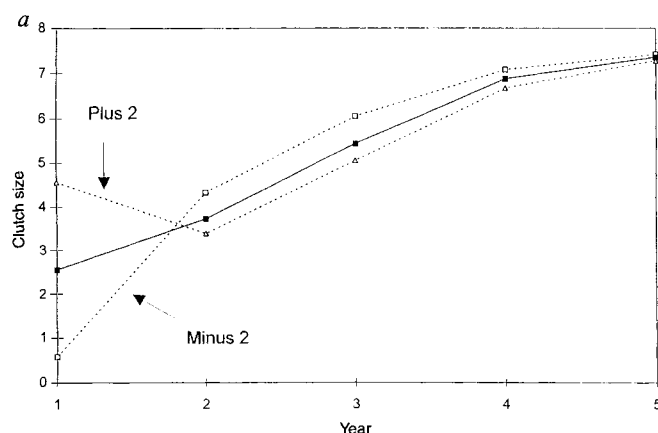
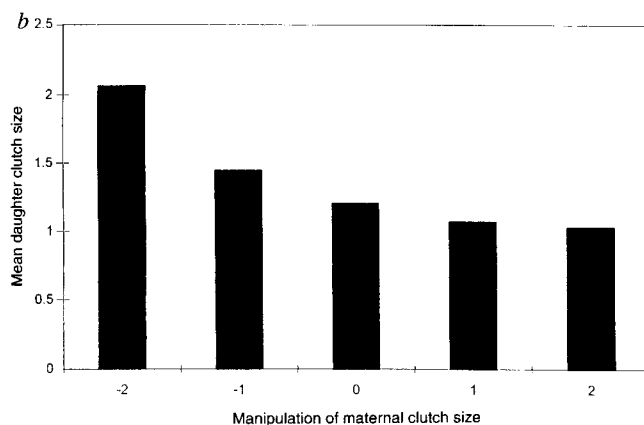


FIG. 3 Effect of manipulation of maternal clutch size when reproductive effort affects offspring survival and condition and maternal survival and future maternal condition. The manipulation is performed in year 1 on a mother in condition $x = 0.625$. a, Subsequent maternal mean clutch sizes.



b, Clutch size at age 1 of daughters from the manipulated clutch. Model 1 modified so that offspring condition is $0.5(r^2/(r^2 + 0.4^2))$. $R = 4ux$, $m_s = 0.75$, $m_e = 0.75$ and $\alpha = 0.1596$.

As we have already said, counting offspring is clearly an inadequate measure when offspring differ in quality, but even counting grandchildren may not be adequate³³. If intergenerational effects are very strong, then the qualities of grandchildren may depend strongly on maternal condition and behaviour. This means that all grandchildren are not equivalent, and so counting them does not measure fitness.

Trivers and Willard⁵⁸ considered whether a female should produce sons or daughters. They assume that males gain relatively more than females from being of high quality and that there are maternal effects so that high-quality females produce high-quality offspring. From this they conclude that females of good quality should prefer to produce sons rather than daughters. Their argument measures the value of sons and daughters solely in terms of the number of offspring that they can produce and fails to take into account the reproductive values of these offspring. Leimar⁵⁹ shows that when the contribution of future maternal effects on reproductive value is taken into account, the reproductive value of a high-quality daughter may be greater than the reproductive value of a high-quality son. Under these circumstances, high-quality females should produce daughters rather than sons.

Manipulations

One of the major difficulties in using a state-dependent approach to life histories is that it may not be clear what are the important components of an organism's state. Even once the key state variables have been identified, the trade-offs (as represented by functions f_1 to f_4) need to be established. It is widely recognized that trade-offs cannot be inferred from phenotypic correlations^{14,60}; indeed positive correlations between components of fitness can be seen as evidence of underlying differences in quality. In birds, many workers have investigated trade-offs by manipulating clutch size. The work of Gustafsson¹⁰ on the collared flycatcher suggests that effort decreases a condition variable. The problem is that the manipulation gives no direct information on the state variable itself, the way in which behaviour changes the state variable, or the way in which the state variable determines behaviour. By manipulating the model, we can hope to gain insight into how these hidden features determine the effects that a manipulation will have, and hence infer something about these features from the observed results of manipulations. Figure 3 illustrates the result of manipulation of maternal clutch size predicted by a modified version of model 1. Subsequent maternal clutch size may be influenced by the manipulation for several years (Fig. 3a). The manipulation of a mother's clutch size can have

an effect on the clutch size of daughters from the manipulated clutch (Fig. 3b). This figure shows that a relatively simple state-dependent model can capture many of the effects observed by Gustafsson¹⁰.

Maternal rank inheritance

In many groups of animals, individuals differ in their ability to gain access to resources. This ability is usually characterized by an individual's rank or status. In many primate species and in spotted hyenas, a female's status is correlated with her mother's status⁶¹. This feature is known as maternal rank inheritance. In several species, there is known to be a positive correlation between status and reproductive success⁶². We now present a schematic model (model 3) of the effects of status, which, while not providing a detailed representation of any particular species, illustrates the applicability of the state-dependent approach to social systems.

Model 3. In this model the state variable represents the animal's social status, which may change from year to year. If the animal puts no effort into maintaining its status, then its status will decrease. An animal may maintain or increase its status by the expenditure of effort, but an increase in effort reduces its annual survival. We envisage increased effort to include increased aggressive

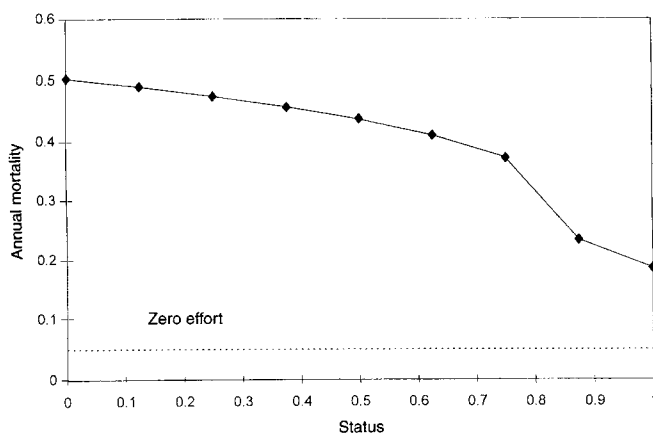


FIG. 4 Annual mortality as a function of status in the model of maternal rank inheritance (model 3). In this model the female can raise her status at the cost of increasing her probability of being killed. The dashed line shows the mortality level when the female makes no attempt to maintain status ($u = 0$). $\alpha = 0.6106$.

behaviour, so the increase in mortality may be a consequence of injuries sustained while fighting. There are no reproductive decisions. Each female always produces one offspring each year, and this offspring survives to maturity with a probability that is proportional to its mother's status. The status of the offspring at age 1 is equal to the status of the mother when the offspring was born. Details of functions f_1 to f_4 for this model are given in Table 1.

A life-history strategy specifies the level of effort devoted to maintaining or raising status. Raising status has two potential advantages: the survival probability of offspring is increased, and any surviving offspring will have higher status. To achieve these advantages, the optimal life-history strategy results in a high level of mortality (Fig. 4). This figure is based on a potentially long-lived species with a mean lifetime of 20 years if no effort is expended. Mortality is lowest for high-status individuals but even they expect to live for less than 6 years.

The presence of maternal effects greatly increases the reproductive value of high-status individuals relative to low-status individuals. For example, with the parameter values given in Fig. 4, a mother with status 1.0 produces twice as many offspring as a mother with status 0.5, but leaves about 30 times as many descendants in the distant future. This means that even when the population is large, it is likely to be dominated by a few maternal lines. This may have important consequences for the diversity in mitochondrial DNA observed in populations.

Model 3 is of course schematic; it does not provide a detailed model of social status. Perhaps its most important shortcoming is that it does not represent status as a relationship between members of a group. Status is really a relative measure, and needs to be analysed by game-theoretic models. A representation of a particular social system would also have to take into account the appropriate biological details. Status can have a variety of effects on components of reproductive success. In model 3, status influences offspring survival; this effect is found in the rhesus macaque⁶³. In savannah baboons, daughters of high-ranking females become sexually mature sooner than daughters of low-ranking females⁶⁴. In the spotted hyena, sons of the highest-ranking female are able to feed in the presence of adult females and can remain in the natal clan for longer than sons of low-ranking females⁶⁵. Consequently, the sons of the highest-ranking female are likely to have greater reproductive success than the sons of lower-ranking females. In all of these cases it is not sufficient to assess reproductive success solely in terms of the number of offspring produced.

Phenotypic plasticity

The phenotype produced by a given genotype may depend on the environment in which it develops. This is referred to as phenotypic plasticity, and the relationship between the environment and the phenotype is known as the genotype's norm of reaction. In some cases, this reaction of a genotype to its environment may be adaptive. Variation in size and age at maturity as a function of conditions during growth has been considered in this way^{66,67}. Within life-history theory, a genotype is regarded as coding an organism's life-history strategy. In state-dependent life-history theory, a strategy is a rule which specifies how an organism should respond to its internal state and its environment. Within this framework, strategies are allowed to be plastic, thus state-dependent life-history theory provides a natural setting in which to discuss adaptive plasticity³⁸.

In some cases the optimal norm of reaction is relatively easy to find. When all individuals in a habitat are in the same state at birth and habitats are isolated from one another, then the optimal strategy is to maximize the growth rate of descendant numbers in each habitat. When there is density-dependent regulation of population size and all newborn individuals are in the same state, each individual should maximize the expected number of offspring produced over its lifetime⁶⁸.

Other cases (for example, when animals may migrate between

habitats) may require a more complex analysis^{38,69}. Fitness is a property of strategies, not of individuals or individual actions, and in general the best action to perform in one state depends on what the organism would have done if it had been in another state^{37,38}. This is because descendants may be found in states other than the mother's current state. The value of these descendants depends on the actions performed by them, and the rule for actions is inherited from the mother. The take-home message is that natural selection should be thought of as acting on an organism's norm of reaction, not on its behaviour in a particular state.

Limitations of the framework

The modelling framework that we have presented ignores some important biological features which we now discuss.

Annual routines. In the formalism that we have presented, reproductive decisions are made at yearly intervals. In practice, many organisms make a sequence of state-dependent life-history decisions through the year. The formalism is readily extendible to include this case. Under an optimal strategy, the sequence of decisions through the year maximizes the expected reproductive value of descendants at the end of the year, and may be found by dynamic programming.

Frequency dependence. In general, we expect organisms to be subject to frequency-dependent and density-dependent effects. We then expect a population to evolve to an evolutionarily stable strategy (ESS). A population adopting a given strategy collectively creates an environment for members of the population. Suppose that almost all members of a population adopt a given strategy, and hence create a particular environment. Then this resident population strategy is an ESS if mutants adopting different strategies do worse in this environment than members adopting the population strategy.

There are thus two components to finding an ESS: (1) determining the environment created by a resident strategy, and (2) determining the optimal strategy given this environment, and hence establishing whether any mutant can do better than the resident population. The state-dependent framework that we have presented tackles the second component, and is thus a necessary part of a full ESS analysis.

Fluctuating environments. When there are year to year variations in factors such as weather and population density, then the measure of fitness that we have used here is no longer adequate. In a non-fluctuating environment, each individual maximizes its expected number of descendants left far into the future. There is no such individual-based optimization criterion in a fluctuating environment^{70,71}. When the population is structured by some state variable and the environment is fluctuating, the analysis becomes very much more complex than in the non-fluctuating case, but some results have been established^{72,73}. □

John M. McNamara is at the School of Mathematics, University of Bristol, Bristol BS8 1TW, UK. Alasdair I. Houston is at the School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK.

1. Roff, D. A. *The Evolution of Life Histories* (Chapman and Hall, New York, 1992).
2. Stearns, S. C. *The Evolution of Life Histories* (Oxford Univ. Press, 1992).
3. Pianka, E. R. & Parker, W. S. *Am. Nat.* **109**, 453–464 (1975).
4. Clutton-Brock, T. H., Guinness, F. E. & Albon, S. D. *Red Deer: Behavior and Ecology of Two Sexes* (University of Chicago Press, 1982).
5. Pettifor, R. A., Perrins, C. M. & McCleery, R. H. *Nature* **336**, 160–162 (1988).
6. Daan, S., Dijkstra, C. & Tinbergen, J. M. *Behaviour* **114**, 83–116 (1990).
7. Weimerskirch, H. *Oikos* **64**, 464–473 (1992).
8. Lunn, N. J. & Boyd, I. L. *J. Zool.* **229**, 55–67 (1993).
9. Risch, T. S., Dobson, F. S. & Murie, J. O. *Ecology* **76**, 1643–1654 (1995).
10. Gustafsson, L. et al. *Phil. Trans. R. Soc. Lond. B* **346**, 323–331 (1994).
11. Perrins, C. M. & Moss, D. J. *Anim. Ecol.* **44**, 695–706 (1975).
12. Pettifor, R. A. *J. Anim. Ecol.* **62**, 131–144 (1993).
13. Boyce, M. S. & Perrins, C. M. *Ecology* **68**, 142–153 (1987).
14. van Noordwijk, A. J. & de Jong, G. *Am. Nat.* **128**, 137–142 (1986).
15. Smith, C. C. & Fretwell, S. D. *Am. Nat.* **108**, 499–506 (1974).
16. Lloyd, D. G. *Am. Nat.* **129**, 800–817 (1987).
17. Lande, R. & Kirkpatrick, M. *Genet. Res.* **55**, 189–197 (1990).
18. Perrins, C. M. *J. Anim. Ecol.* **34**, 601–647 (1965).
19. Witter, M. S. & Cuthill, I. C. *Phil. Trans. R. Soc. Lond. B* **340**, 73–92 (1993).
20. Houston, D. C. et al. *Ibis* **137**, 322–328 (1995).
21. Heinsohn, R. G. *Am. Nat.* **137**, 864–888 (1991).

22. Norris, K., Anwar, M. & Read, A. F. J. *Anim. Ecol.* **63**, 601–610 (1994).
23. Møller, A. P. J. *Anim. Ecol.* **62**, 309–322 (1993).
24. Clutton-Brock, T. H., Albon, S. D. & Guinness, F. E. in *Reproductive Success* (ed. Clutton-Brock, T. H.) 325–343 (University of Chicago Press, 1988).
25. Perrin, N., Sibily, R. M. & Nichols, N. K. *Evol. Ecol.* **7**, 576–592 (1993).
26. Lefkovich, L. P. *Biometrics* **21**, 1–18 (1965).
27. Caswell, H. *Matrix Population Models* (Sinauer, Sunderland, MA, 1989).
28. Tanner, J. E., Hughes, T. P. & Connell, J. H. *Ecology* **75**, 2204–2219 (1994).
29. Brault, S. & Caswell, H. *Ecology* **74**, 1444–1454 (1993).
30. Houston, A. I. et al. *Nature* **332**, 29–34 (1988).
31. McNamara, J. M. *Theor. Pop. Biol.* **40**, 230–245 (1991).
32. McNamara, J. M. *Acta biotheor.* **41**, 165–174 (1992).
33. McNamara, J. M. & Houston, A. I. *Evol. Ecol.* **6**, 170–185 (1992).
34. Mangel, M., Rosenheim, J. A. & Adler, F. A. *Behav. Ecol.* **5**, 412–417 (1994).
35. Charlesworth, B. *Evolution in Age-structured Populations* 2nd edn (Cambridge Univ. Press, 1994).
36. Fisher, R. A. *The Genetical Theory of Natural Selection* 2nd edn (Dover, New York, 1958).
37. McNamara, J. M. *J. theor. Biol.* **161**, 23–37 (1993).
38. Houston, A. I. & McNamara, J. M. *Evol. Ecol.* **6**, 243–253 (1992).
39. Marrow, P. et al. *Phil. Trans. R. Soc. Lond. B* **351**, 17–32 (1996).
40. Cochran, M. E. & Ellner, S. *Ecol. Monogr.* **62**, 345–364 (1992).
41. Saether, B.-E. & Heim, M. J. *Anim. Ecol.* **62**, 482–489 (1993).
42. Green, W. C. H. & Rothstein, A. *Oecologia* **86**, 521–527 (1991).
43. Reiter, J. & Le Boeuf, B. J. *Behav. Ecol. Sociobiol.* **28**, 153–160 (1991).
44. Clutton-Brock, T. H., Major, M. & Guinness, F. E. *J. Anim. Ecol.* **54**, 831–846 (1985).
45. Charlesworth, B. & Leon, J. A. *Am. Nat.* **110**, 449–459 (1976).
46. Clutton-Brock, T. H. *Am. Nat.* **123**, 212–229 (1984).
47. Partridge, L. & Barton, N. H. *Nature* **362**, 305–311 (1993).
48. Blarer, A., Doebeli, M. & Stearns, S. C. *Proc. R. Soc. Lond. B* **262**, 303–312 (1995).
49. Kirkwood, T. B. L. & Rose, M. R. *Phil. Trans. R. Soc. Lond. B* **332**, 15–24 (1991).
50. Prince, P. A., Ricketts, C. & Thomas, G. *Condor* **83**, 238–242 (1981).
51. Chandra, R. K. *Science* **190**, 289–290 (1975).
52. Zamenhof, S., van Marthens, E. & Grauel, L. *Science* **172**, 850–851 (1971).
53. Mech, L. D., Nelson, M. E. & McRoberts, R. E. *J. Mamm.* **72**, 146–151 (1991).
54. Lumey, L. H. *Paediatr. Perinat. Epidemiol.* **6**, 240–253 (1992).
55. Price, K. & Boutin, S. *Behav. Ecol.* **4**, 144–150 (1993).
56. Mousseau, T. A. & Dingle, H. A. *Rev. Ent.* **36**, 511–534 (1991).
57. Roach, D. A. & Wulff, R. D. A. *Rev. Ecol. Syst.* **18**, 209–235 (1987).
58. Trivers, R. L. & Willard, D. E. *Science* **179**, 90–92 (1973).
59. Leimar, O. *Behav. Ecol.* (in the press).
60. Partridge, L. & Harvey, P. H. *Science* **241**, 1449–1455 (1988).
61. Holekamp, K. E. & Smale, L. *Anim. Behav.* **46**, 451–466 (1993).
62. Harcourt, A. H. *Trends Ecol. Evol.* **4**, 101–105 (1989).
63. Silk, J. B. *Am. Nat.* **121**, 56–66 (1983).
64. Aitmann, J., Hausfater, G. & Aitmann, S. A. in *Reproductive Success* (ed. Clutton-Brock, T. H.) 403–418 (University of Chicago Press, 1988).
65. Frank, L. G. *Anim. Behav.* **34**, 1510–1527 (1986).
66. Stearns, S. C. & Koella, J. C. *Evolution* **40**, 893–913 (1986).
67. Reznick, D. N. *J. evol. Biol.* **3**, 185–203 (1990).
68. Bernigan, D. & Koella, J. C. *J. evol. Biol.* **7**, 549–566 (1994).
69. Kaweck, T. J. & Stearns, S. C. *Evol. Ecol.* **7**, 155–174 (1993).
70. Ellner, S. J. *theor. Biol.* **123**, 173–185 (1986).
71. McNamara, J. M. *Evol. Ecol.* **9**, 185–203 (1995).
72. Metz, J. A. J. et al. *Trends Ecol. Evol.* **7**, 198–202 (1992).
73. McNamara, J. M. *Theor. Pop. Biol.* (in the press).

ACKNOWLEDGEMENTS. We thank I. Boyd, T. Clutton-Brock, I. Cuthill, P. Harvey, J. Hutchinson, O. Leimar, L. Partridge, R. Pettifor, A. Sih, S. Stearns, S. Verhulst and M. Witter for comments on previous versions of this paper.

S-nitrosohaemoglobin: a dynamic activity of blood involved in vascular control

Li Jia*, Celia Bonaventura†‡, Joseph Bonaventura†‡ & Jonathan S. Stamler*†

* Department of Medicine, Divisions of Respiratory and Cardiovascular Medicine and † Department of Cell Biology, Duke University Medical Center, MSRB Building, Room 321, Box 2612, Durham, North Carolina 27710, USA

‡ Nicholas School of the Environment, Duke Marine Biomedical Center, Pivers Island, North Carolina 28516, USA

A dynamic cycle exists in which haemoglobin is S-nitrosylated in the lung when red blood cells are oxygenated, and the NO group is released during arterial–venous transit. The vasoactivity of S-nitrosohaemoglobin is promoted by the erythrocytic export of S-nitrosothiols. These findings highlight newly discovered allosteric and electronic properties of haemoglobin that appear to be involved in the control of blood pressure and which may facilitate efficient delivery of oxygen to tissues. The role of S-nitrosohaemoglobin in the transduction of NO-related activities may have therapeutic applications.

INTERACTIONS of haemoglobin (Hb) with small diffusible ligands such as O₂, CO₂ and NO occur at its haem iron and amino termini. O₂/CO₂ delivery in the lung and systemic microvasculature is allosterically controlled. Such control is not known to apply to nitric oxide. Specifically, it is thought that Hb(Fe) is involved in limiting NO's sphere of action^{1,2}, but that NO does not modify the functional properties of Hb to any physiologically significant degree. On the other hand, kinetic modelling predicts that most free NO in the vasculature should be scavenged by Hb¹. Accordingly, the steady-state level of NO may fall below the Michaelis constant (*K_m*) for target enzymes such as guanylate cyclase¹, if not in the unperturbed organism, then with oxidant stress such as that found in atherosclerosis. These considerations raise the fundamental question of how NO exerts its biological activity.

One answer to this paradox may be found in the formation of S-nitrosothiols (RSNOs)³, which retain NO-like vasorelaxant activity⁴, but are not subject to the diffusional constraints imposed by the high concentration of Hb in the blood. The NO group of RSNOs possesses nitrosonium (NO⁺) character that distinguishes

it from NO itself, and enables it to elicit responses of which nitric oxide is incapable^{5,6}. Moreover, -SNO groups in proteins may serve a signalling function, perhaps analogous to phosphorylation^{4,6}. Although S-nitrosylation can regulate protein function^{4,6}, intracellular S-nitrosothiols—the *sine qua non* of a regulatory post-translational modification—have not so far been demonstrated.

Haemoglobin is a tetramer composed of two α - and two β -subunits. In human Hb, each subunit contains one haem, and the β -subunits also contain highly reactive SH groups (Cys β 93) (refs 7,8). These cysteine residues are highly conserved among species, although their function has remained unknown. We therefore explored the possibility that S-nitrosohaemoglobin (SNO-Hb) forms within red blood cells and serves a regulatory function.

Interactions of NO and RSNO with Hb

Naturally occurring N-oxides such as NO and RSNOs^{3,9,10} differed markedly in their reactions with Hb. NO bound very rapidly to deoxyHb (Hb[FeII]), forming relatively stable Hb[FeII]NO

complexes (Fig. 1a), and converted oxyHb (Hb[Fe(II)O₂]) to methaemoglobin (Hb[Fe(III)]) and nitrate (Fig. 1b), confirming previous reports^{7,11}. In contrast, RSNOs did not react with the haems of either deoxyHb or Hb[Fe(II)O₂] (Fig. 1c,d). Rather, they participated in transnitrosation reactions with the sulphhydryl groups of Hb as described in the legends of figures 1c and d.

Allosteric control of S-nitrosylation

Oxygenation of Hb is associated with an allosteric transition that increases the reactivity of Cys β 93 to alkylating reagents^{12–14}, the physiological importance of which was never established. We observed that rates of S-nitrosylation of Hb were markedly dependent on conformational state. In the oxy conformation (R state), S-nitrosylation was faster than in the deoxy conformation (T state) (Fig. 2a). The rate of S-nitrosylation was accelerated in both conformations by alkaline conditions (rate at pH 9.2 > pH 7.4), which exposes the Cys β 93 that is otherwise screened by the C-terminal His146 β because the salt bridge (Asp β 94–His β 146) tying down the histidine is loosened at high pH. These data suggest that the increase in thiol reactivity associated with the R state derives, at least in part, from improved NO access rather than a conformation-induced change in pK.

NO/RSNO interactions in RBCs

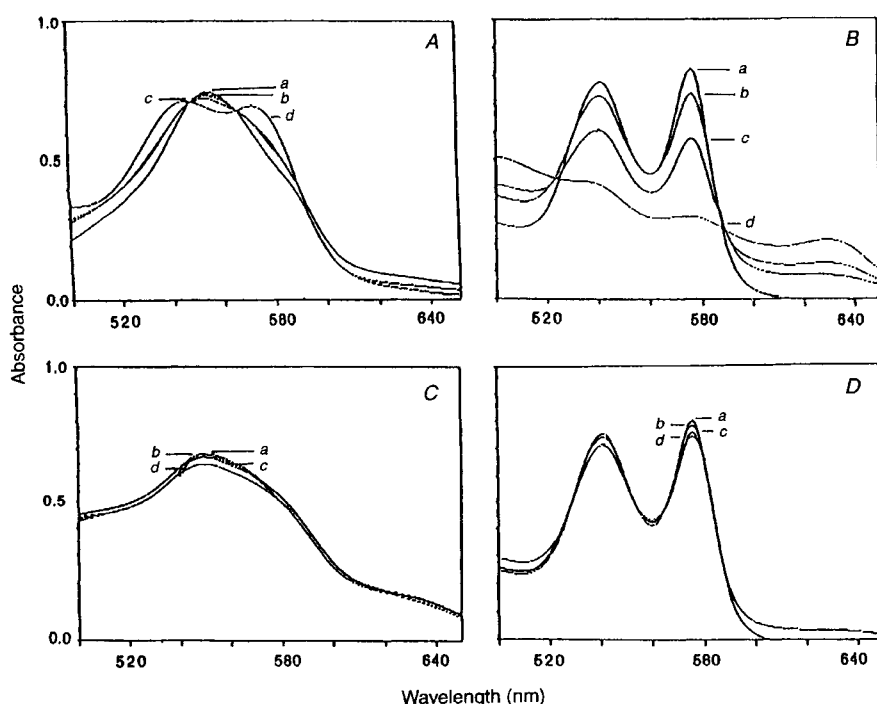
Given that Hb is largely contained within red blood cells (RBCs), we explored potential mechanisms for S-nitrosylation of intracellular protein. Incubation of oxygenated rat RBCs with S-nitrosocysteine resulted in very rapid formation of intra-

cellular SNO-Hb[Fe(II)O₂] (Fig. 3a). S-nitrosylation was faster than oxidation of Hb under these conditions. Intraerythrocytic SNO-Hb also formed when RBCs were treated with S-nitrosohomocysteine or S-nitrosocysteinylglycine (data not shown), but not with S-nitrosogluthathione (GSNO). Thus, erythrocyte access of RSNOs is thiol-group specific. Exposure of oxygenated RBCs to NO produced mainly metHb.

EDRF and Hb

Hb-mediated inhibition of endothelium-dependent relaxations is commonly used as a marker of NO responses. Inasmuch as reactions with either metal or thiol centres of Hb should attenuate NO/EDRF (endothelium-derived relaxing factor) responses, we investigated the molecular basis of inhibition. Hb preparations in which β 93 thiol groups had been blocked with N-ethylmaleimide (NEM) or the haems blocked by cyanmet (Fe(CN)) derivitization were studied in an aortic ring bioassay, and their activities compared with that of native Hb. Both cyanmet-Hb and NEM-Hb caused an increase in vessel tone and attenuated acetylcholine (EDRF)-mediated relaxations (Fig. 3b). However, native Hb was significantly more effective than either of the modified Hb preparations (Fig. 3b). Together these results suggest that both the thiol and metal groups of Hb contribute to its NO-related activity. To verify formation of an S-nitrosothiol in Hb, we established a bioassay in which 2-cm segments of thoracic aorta were interposed in Tygon tubing, through which 3 ml of Krebs solution containing Hb (4 μ M) and ACh (2 μ M) were circulated by roller pump (1.5 ml min⁻¹ $\times$ 5 min). Analysis of the

FIG. 1 Interactions of NO and RSNO with Hb. A, Interaction of NO with deoxyHb. Conversion of deoxyHb (Hb[Fe(II)]) to Hb[Fe(II)NO] is observed upon incubation of Hb[Fe(II)] with increasing concentrations of nitric oxide. a, Deoxy Hb; b–d, reaction mixtures of NO and Hb[Fe(II)] in ratios of 1:1, 2:1 and 10:1, respectively. The reaction product Hb[Fe(II)NO], formed essentially instantaneously on addition of NO (that is, within instrument dead time). B, Interaction of NO with oxyHb. Conversion of oxyHb (Hb[Fe(II)O₂]) to metHb (Hb[Fe(III)]) is observed upon incubation of oxyHb with increasing concentrations of NO. a, Oxy Hb; b–d, reaction mixtures containing NO and oxyHb in ratios of 1:1, 2:1 and 10:1, respectively. Methaemoglobin formation occurred instantaneously on addition of NO (within instrument dead time). C, Interaction of S-nitrosothiols with deoxyHb. There is little (if any) interaction of RSNO with the haem of Hb. On the other hand, conversion of Hb[Fe(II)] to SNO-Hb[Fe(II)] is observed upon incubation of either GSNO (shown) or S-nitrosocysteine (CYSNO) with deoxyHb (see below). a, DeoxyHb; b–d, reaction mixtures of GSNO and Hb[Fe(II)] in ratios of 1:1, 2:1 and 10:1, respectively. Spectra were taken after 60 min of incubation in b, c, and 15 min in d. Further analysis of reaction products revealed the formation of moderate amounts of SNO-Hb in all cases. Yields of SNO-Hb (S-NO/Hb) in b, c and d at 60 min were 2.5, 5 and 18.5%, respectively. See more on SNO-Hb formation in D and Fig. 2a. D, Interaction of S-nitrosothiols with oxyHb. There is little (if any) reaction of GSNO (or CYSNO) at the haem centres of Hb[Fe(II)O₂]. Specifically, the capacity of O₂ binding to haem is unaffected by RSNOs. On the other hand, conversion of Hb[Fe(II)O₂] to SNO-Hb[Fe(II)O₂] is observed upon incubation of either GSNO (shown) or CYSNO with oxyHb (see below). a, OxyHb; b–d, reaction mixtures of GSNO and oxyHb in ratios of 1:1, 2:1 and 10:1, respectively. Spectra were taken after 60 min of incubation in the spectrophotometer. Further analysis of reaction products revealed the formation of SNO-Hb in all cases. Yields of SNO-Hb in spectra b, c and d were 5, 10 and 50% (S-NO/Hb), respectively. In 5 other determinations, the yield of S-NO/Hb was 0.37 $\pm$ 0.06 using GSNO (pH 7.4; 10-fold excess over Hb) and $\sim$ 2 S-NO/tetramer (1.97 $\pm$ 0.06) using CYSNO. These last data are in agreement with reports that human HbA contains 2 titratable SH groups.



METHODS. Human HbA₀ was purified from red cells as described²⁴. Nitric oxide solutions were rigorously degassed and purified according to standard procedures²⁵ and saturated solutions were transferred in air-tight syringes. Deoxygenation of Hb was achieved by addition of excess dithionite (NO studies) or by reduction of Hb[Fe(II)O₂] through evacuation in Thunberg tubes (RSNO studies; as RSNOs react with dithionite). RSNOs were synthesized as described^{3,16}. Incubations with HbA₀ were made in phosphate buffer, pH 7.4, 0.5 mM EDTA. Quantifications of SNO-Hb were made according to the method of Saville^{3,4} after purification of protein with Sephadex G-25 columns. No low-molecular-weight S-NO complexes survived this purification and all activity was protein-precipitable. Reactions and spectra were run in a Perkin Elmer UV/Vis Spectrometer, Lambda 2S.

effluent (see Table 1) revealed the formation of SNO-Hb (20 ± 4 nM) in 5 of 5 experiments.

Circulatory dynamics

To determine whether SNO-Hb is naturally occurring in the blood and, if so, what its relationship is to the O_2 transport capacity and nitrosylated-haem content of RBCs, we developed an analytical approach to assay the *S*-nitrosothiol and nitrosyl-haem content of erythrocytes (Table 1). Arterial blood was obtained from the left ventricle of anaesthetized rats by direct puncture; venous blood was obtained from the jugular vein and inferior vena cava. Hb was then purified from red cells and assayed for RSNO and (FeII)NO content. Interestingly, we found that arterial blood contained significant levels of SNO-Hb, whereas levels were virtually undetectable in venous blood (Table 1). Measurements made 45 min after infusion of the NO synthase inhibitor N^G -monomethyl-L-arginine (L-NMMA) (50 mg kg^{-1}) showed a depletion of SNO-Hb as well as total Hb-NO (82 and $50 \pm 18\%$, respectively; $n = 3-5$; $P < 0.05$). These results establish the endogenous origin of SNO-Hb, although some environmental contribution is not excluded. The arterial-venous distribution of SNO-Hb was reversed in the case of Hb(FeII)NO, which was detected in higher concentrations in partially deoxygenated (venous) erythrocytes (Table 1). Accordingly, the proportion of nitrosylated protein thiol and haem appears to depend on the oxygenation state of the blood. Consistent with our findings, others have shown that Hb(FeII)NO forms mainly in venous (partially deoxygenated) blood¹⁵. However, levels of Hb(FeII)NO *in vivo* are typically too low to be detected (by electron paramagnetic resonance, EPR) and SNO-Hb is EPR-silent (it is not paramagnetic). Photolysis-chemiluminescence is the first method capable of accurately measuring NO binding to Hb under physiological conditions.

Bioactivity and regulation of SNO-Hb

Arterial RBCs contain two physiologically important forms of haemoglobin: Hb(FeII) O_2 and Hb(FeIII) (ref. 8). Arterial-venous differences in the *S*-nitrosothiol content of intraerythrocytic Hb suggest that the NO group is released during red cell transit. This raises the possibility of functional consequences, perhaps influenced by the redox state of haem and its occupation by ligand. SNO-Hb(FeII) O_2 has modest NO-like activity when tested in a vascular ring bioassay. Specifically, the contraction elicited by SNO-Hb(FeII) O_2 is less than that of native Hb(FeII) O_2 , indicating that *S*-nitrosylation partially reverses the contractile effects of Hb (Fig. 4a). On the other hand, SNO-Hb(FeIII) is a vasodilator (Fig. 4a). Notably, free NO has no relaxant activity in the presence of Hb(FeII) O_2 or Hb(FeIII) (results not shown).

Red blood cells contain millimolar concentrations of glutathione (GSH). As equilibria among RSNOs are rapidly established through RSNO/thiol exchange¹⁶, the vasoactivity of SNO-

TABLE 1 Endogenous levels of *S*-nitrosohaemoglobin and nitrosyl(FeII)-haemoglobin

Blood	SNO-Hb (nM)	Hb(FeII)NO (nM)
Arterial	$311 \pm 55^*$	$536 \pm 99^\dagger$
Venous	32 ± 14	894 ± 126

Blood was obtained from the left ventricle (arterial) and jugular vein (venous) of anaesthetized Sprague-Dawley rats. Comparable venous values were obtained in blood from the inferior vena cava. Red blood cells were isolated by centrifugation at 800g, washed three times in PBS at 4 °C, lysed by the addition of 4-fold excess volume of deionized water containing 0.5 mM EDTA, and desalted rapidly across G-25 columns at 4 °C. Hb samples from 24 rats were divided into two aliquots which were then treated or not treated with 10-fold excess $HgCl_2$ over protein ($HgCl_2$ destroys RSNOs with stoichiometric production of nitrite). Determinations of SNO-Hb and Hb(FeII)NO were made by photolysis-chemiluminescence according to principles described below. In 12 additional rats, SNO-Hb was assayed from nitrite determination after $HgCl_2$ treatment. Specifically, samples (with and without $HgCl_2$) were separated on Amicon-3 centrifron filters (M_w cut-off 3,000) at 4 °C for 1 h, and the low-molecular-weight fractions collected in air-tight syringes containing $1 \mu M$ glutathione in 0.5 N HCl. Under these conditions, any nitrite present was converted to *S*-nitrosoglutathione, which was then measured by photolysis-chemiluminescence (detection limit ~ 1 nM). SNO-Hb was present in all arterial samples, and amounts determined by this method (286 ± 33 nM) were virtually identical to and not statistically different from those shown in the table. In venous blood, SNO-Hb was undetectable (0.00 ± 25 nM); levels were not statistically different from those given. Highly sensitive photolysis-chemiluminescence methodology, originally used to measure RSNOs in biological systems^{3,19}, involves photolytic liberation of NO from thiol, which is then detected in a chemiluminescence spectrometer by reaction with ozone. The same principle of operation can be used to cleave (and measure) NO from nitrosyl-metal compounds⁹. With adjustment of flow rates in the photolysis cell, we verified that complete photolysis of the NO ligand of Hb(FeII)NO could be achieved. Standard curves derived from synthetic preparations of SNO-Hb, Hb(FeII)NO, and *S*-nitrosoglutathione were linear ($R < 0.99$), virtually superimposable, and had sensitivity limits of ~ 1 nM. Two analytical criteria reliably distinguished SNO-Hb from Hb(FeII)NO: (1) signals from SNO-Hb were eliminated by pretreatment of samples with 10-fold excess $HgCl_2$, whereas Hb(FeII)NO was resistant to mercury challenge; and (2) treatment of SNO-Hb with $HgCl_2$ produced nitrite (by standard Griess reactions) in quantitative yield, whereas similar treatment of Hb(FeII)NO did not. UV/visible spectroscopy confirmed that NO remained attached to haem in the presence of excess $HgCl_2$.

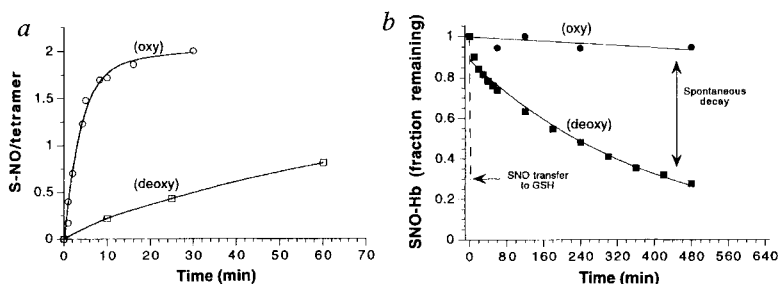
* $P < 0.05$ versus venous.

† $P < 0.05$ for paired samples versus venous.

Hb was reassessed in the presence of glutathione. Figure 4b shows that glutathione potentiates the vasodilator activity of both SNO-Hb(FeII) O_2 and SNO-Hb(FeIII). *S*-nitrosoglutathione (GSNO) formation (confirmed chemically and by bioassay) appeared to account for this effect. Further kinetic analysis revealed that more

FIG. 2 Allosteric function of O_2 in regulation of Hb *S*-nitrosylation. a, Oxygenation accelerates *S*-nitrosylation of Hb. Rates of Hb *S*-nitrosylation by *S*-nitrosocysteine (CYSNO) are faster in the oxy conformation (Hb(FeII) O_2) than in the deoxy state (Hb(FeII)). b, Deoxygenation accelerates denitrosylation of Hb. Rates of RSNO decomposition (and transfer) are much faster in the deoxy conformation [SNO-Hb(FeII)] than in the oxy state [SNO-Hb(FeII) O_2]. The decomposition of SNO-Hb(FeII) is further accelerated by the presence of excess glutathione. Within the dead time of our measurements (~ 15 s), a major fraction of SNO-Hb(FeII) was converted to GSNO (dashed line).

METHODS. a, Incubations were performed using 10-fold excess CYSNO over protein ($50 \mu M$) in aerated 2% borate, 0.5 mM EDTA (oxyHb), or in a tonometer after rapid O_2 evacuation (deoxyHb). At the indicated times, samples were rapidly desalted across G-25 columns (pre-equilibrated with PBS, 0.5 mM EDTA, pH 7.4) to remove CYSNO, and analysed for SNO-Hb. b, Hbs in PBS (0.5 mM EDTA, pH 7.4) were incubated in air (oxy) or in a tonometer evacuated of O_2 (deoxy). SNO-Hb(FeII) O_2 decomposition was determined by the method of Saville⁴. Spontaneous decomposition of SNO-Hb(FeII) was followed spectrophotometrically by formation of Hb(FeII)NO. Transnitrosation reactions with glutathione were performed by addition of 100-fold excess glutathione over protein ($50 \mu M$), immediate processing of the reaction mixture under anaerobic conditions, followed by rapid precipitation with trichloroacetic acid (TCA), and analysis of RSNO in the supernatant. Rates of NO group transfer were too rapid to measure accurately by the standard methods used here.



transnitrosation of glutathione occurred in the equilibrium with SNO-Hb(FeII) than SNO-Hb(FeII)O₂ (Fig. 4c). Given the findings of steady-state levels of SNO-Hb in red blood cells (Table 1 and Fig. 3a), these results suggest that (1) the equilibrium between naturally occurring RSNOs and Hb(cys β 93) lies towards SNO-Hb under physiological conditions; (2) transnitrosation reactions involving SNO-Hb and GSH occur in RBCs (indeed, we have verified the presence of low-molecular-weight RSNOs in erythrocytes loaded with SNO-Hb); and (3) oxidation of the metal centre of Hb will shift the equilibrium toward GSNO, thereby potentiating bioactivity.

We tested for additional mechanisms of NO group release from SNO-Hb. Arterial-venous differences in levels of SNO-Hb raised the possibility that S-NO bond stability may be regulated by the allosteric transition of Hb. Accordingly, we compared the rates of NO group release from SNO-Hb(FeII)O₂ and SNO-Hb(FeII). Deoxygenation accelerated the rate of SNO-Hb decomposition (Fig. 2b). These rates were further accelerated greatly by glutathione in a reaction yielding GSNO (Fig. 2b). Our results illustrate that the reaction of O₂ with haem influences S-NO affinity, and suggest a new allosteric function for Hb.

For SNO-Hb to be of physiological importance it must transduce its NO-related activity across the erythrocyte membrane. We therefore incubated erythrocytes containing SNO-Hb in physiological buffer, and measured the accumulation of extracellular RSNOs. Figure 4d shows that RBCs export low-molecular-weight (trichloroacetic-acid precipitable) S-nitrosothiols under these conditions. Haemolysis in these experiments was trivial (< 0.5%) and correction for lysis did not significantly affect rates of RSNO release. These results establish that an equilibrium exists between

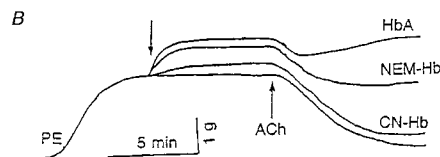
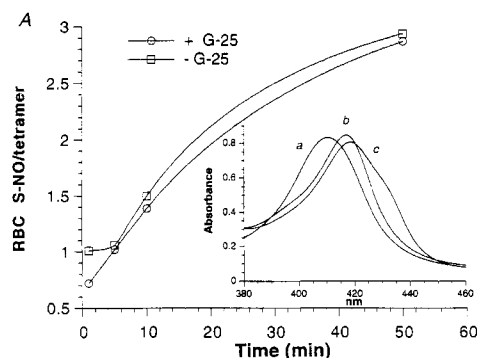
low-molecular-weight and protein RSNOs within the red cell, and that intracellular location is unlikely to be a limiting factor in the transduction of such NO-related activity to the vessel wall.

SNO-Hb/RBCs *in vivo*

Systemic administration of cell-free Hb results in hypertensive responses which have been attributed to NO scavenging by the haem^{17,18}. To determine whether SNO-Hb is free of this adverse effect and whether *in vitro* mechanisms of NO release are valid *in vivo*, we compared responses to Hb and SNO-Hb infused as a bolus into the femoral vein of anaesthetized rats. As shown in Fig. 5, Hb(FeII)O₂ (200 nmol kg⁻¹) caused an increase in mean arterial pressure of 20 ± 3 mm Hg ($n = 4$; $P < 0.05$). In contrast, SNO-Hb(FeII)O₂ did not raise blood pressure and SNO-Hb(FeII) elicited hypotensive responses (Fig. 5). Thus, the behaviour of these compounds *in vivo* closely resembles that seen *in vitro* (Fig. 4a). Moreover, to demonstrate that the responses of red cells are comparable to those of cell-free Hb preparations, erythrocytes containing SNO-Hb were injected into the femoral vein of rats pretreated with 1-NMMA (50 mg kg⁻¹) to deplete endogenous RSNOs. At levels of SNO-Hb comparable to those found in the normal rat (0.1–0.5 μ M), SNO-Hb-containing RBCs elicited hypotensive responses (8 ± 1 mm Hg; mean $\pm$ s.e.m.; $n = 9$), whereas native (SNO-Hb-depleted) RBCs did not ($P = 0.001$). These changes in mean blood pressure of $\sim 10\%$ are of the order of those that differentiate normotension from hypertension in man, and are in the therapeutic range of some antihypertensive regimens. The effects of both Hb and SNO-Hb—whether cell-free or within red cells—were transient, suggesting that S-nitrosylation of Hb and metabolism of SNO-Hb

FIG. 3 NO-related interactions with cysteine residues of Hb in physiological systems. A, S-nitrosylation of intraerythrocytic Hb. Incubation of rat erythrocytes with S-nitrosocysteine (equimolar to haem (5 mM); phosphate buffer, pH 7.4, 25 °C) leads to rapid formation of intracellular SNO-Hb(FeII)O₂. MetHb does not form rapidly. Separation of intracellular RSNOs across G-25 columns reveals that only a small percentage exists as low-molecular-weight S-nitrosothiol (for example, GSNO) at most time points. By 60 min, 3 of the 4 available SH groups of Hb are S-nitrosylated (note that rat Hb contains 4 reactive SH groups). Inset shows spectra of SNO-Hb isolated from rat erythrocytes and related analyses. Spectrum a is that of SNO-Hb isolated from erythrocytes following G-25 chromatography. Treatment of a with dithionite results in reduction of the S-NO moiety, liberating free NO which is autocaptured by deoxy Hb, forming Hb(FeII)NO (note that dithionite simultaneously deoxygenates Hb) (spectrum c). This spectrum (c) reveals a stoichiometry of ~ 3 S-NOs per tetramer. The spectrum of Hb(FeII)NO containing 4 NOs per tetramer is shown for comparison (inset, spectrum b). B, Molecular basis of EDRF/Hb interaction. The effects of native Hb on EDRF responses were compared with Hb preparations in which the thiol or haem centres had been blocked by alkylation or cyanmet derivatization, respectively. All preparations of Hb elicited contractions; however, those of native Hb (in which both-SH and metal centres are free for interaction) were most pronounced. Likewise, acetylcholine (ACh)-mediated relaxations were most effectively inhibited by native Hb. Relaxations were inhibited to lesser degrees for cyanmet Hb (CN-Hb) (in which haems were blocked from reaction) and NEM-Hb (in which thiol groups were alkylated by N-ethylmaleimide). These data show that both haem and β 93-SH groups of Hb contribute to reversal of EDRF responses. Direct measurement of SNO-Hb, formed from EDRF and Hb under similar conditions, is described in the text.

METHODS. a, At the intervals shown, red blood cells were pelleted rapidly by centrifugation, washed three times, lysed in deionized water at 4 °C, and the cytosolic fraction rapidly desalted across G-25 columns. Intracellular SNO-Hb was measured^{3,4}, and confirmed spectroscopically (inset) as already described. b, Descending rabbit thoracic aorta were cut into 3 mm rings and mounted on stirrups attached to force transducers (model FT03; Grass Instruments) for measurement of isometric tone. This bioassay system has been described⁴. Cyanmet Hb was prepared from human HbA²⁴. Alkylation of HbA with N-ethylmaleimide was followed by desalting across G-25 sephadex to remove excess NEM. Removal of unmodified Hbcys β 93 was achieved by passage through Hg-containing affinity columns. Alkylation of free SH groups was verified using 5,5'-dithio-bis(2-nitrobenzoic acid).



Additions	% Increase in tension ($\downarrow$)	% ACh relaxation ($\uparrow$)
Hb (1 μ M)	40.8 ± 2.3 ($n = 7$)	31.9 ± 6.9 ($n = 7$)
NEM-Hb (1 μ M)	29.4 ± 1.3 ** ($n = 7$)	60.5 ± 3.9 * ($n = 7$)
CN-Hb (1 μ M)	12.9 ± 3.0 ** ($n = 6$)	80.7 ± 1.0 ** ($n = 4$)
ACh (1 μ M)		98.3 ± 0.6 ($n = 10$)

* $P < 0.001$, compared to Hb; † $P < 0.001$, compared to ACh.

may be occurring *in vivo*, with consequent restoration of blood pressure. The bioactivity of SNO-Hb in blood, where S-NO/haem stoichiometries approach 1:50,000, is a dramatic illustration of the resistance of this NO-related activity to Hb(Fe) inactivation.

Implications

The increase in SNO-Hb content of red cells across the pulmonary circuit (right ventricular inport–left ventricle) suggests that the Hb molecule is S-nitrosylated in the lung. Selective transfer of the NO group from endogenous RSNOs in lung³ and blood⁹ to SH groups of Hb substantiate these findings, but the mechanism(s) of S-nitrosylation *in vivo* is not known. The corresponding decline in Hb(FeII)NO levels across the pulmonary bed reveals a role for the lung either in the elimination of NO or in its intramolecular transfer from haem to Cys93. Together these data extend the list of function-regulating interactions of Hb with small molecules within the respiratory system, previously known to include the elimination of CO and CO₂ and uptake of O₂. As oxygenation of Hb leads to structural changes that increase the NO-related reactivity of Cys93, O₂ may be regarded as an allosteric effector of Hb S-nitrosylation. This is a newly discovered allosteric function for the protein.

The arterial–venous difference in SNO-Hb concentration suggests that the protein acts as an NO group donor in the systemic circulation. Indeed, there is good indication that SNO-Hb functions in regulation of vasomotor tone. In the microcirculation, where blood pressure is controlled, erythrocytes come in intimate contact with endothelial surfaces. Under these conditions, Hb can contract the vasculature by sharply decreasing the

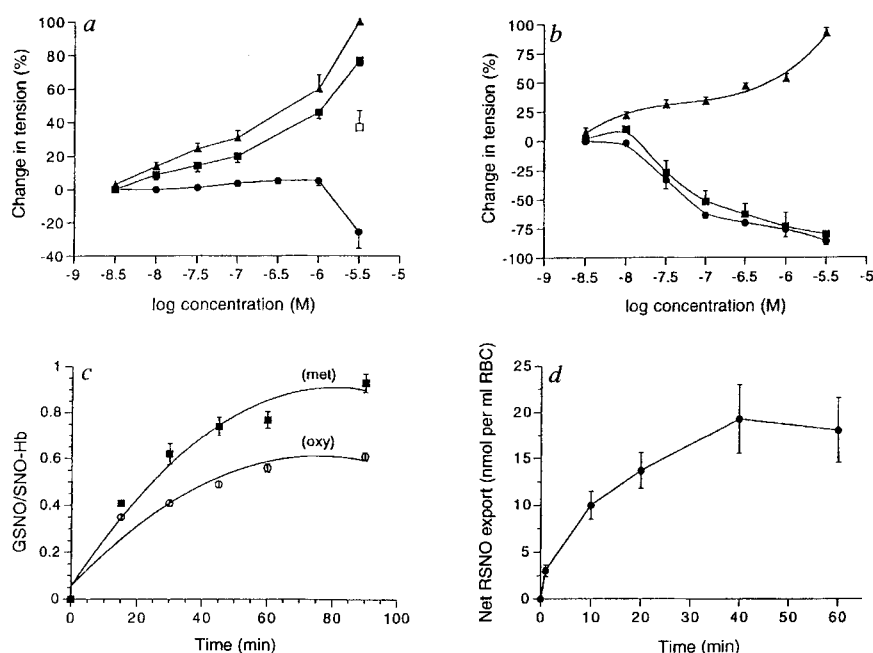
steady-state level of free NO¹. This is believed to contribute to the increase in blood pressure and alterations in blood flow that occur with infusion of cell-free haemoglobins^{17,18}. The transient nature of the hypertensive response, however, is consistent with the subsequent formation of SNO-Hb, as evidenced by its effect of lowering blood pressure at naturally occurring concentrations. Thus, the capacity of the erythrocyte to support the synthesis and metabolism of SNO-Hb may well be important for normal blood flow and delivery of O₂ to tissues.

Mammals must therefore have adopted special molecular mechanisms to ensure adequate delivery of NO in the microcirculation. Our results suggest that Hb may have evolved both electronic and conformational switching mechanisms to achieve NO homeostasis. Specifically, NO scavenging by the metal centre(s) of SNO-Hb(FeII)O₂ would be sensed through its conversion to met(FeII) (Fig. 1b). This electronic switch would effect NO group release (Figs 3a, b and 4a). In this way, the NO-related activity of SNO-Hb would be partly determined by the amount of NO scavenged. The same metal-responsive mechanism in Hb may have been adapted to serve the broader function of sensing cell redox state. Changes in O₂ tension might also regulate NO delivery, as we observed that NO release was facilitated by deoxygenation. This allosteric effect may promote the efficient utilization of O₂ as NO controls mitochondrial respiration²⁶. It is also possible that NO group release serves to regulate capillary blood flow.

S-nitrosothiol groups in proteins take part in NO metabolism and regulation of cellular functions^{4,6}. Our detection of SNO-Hb in erythrocytes is, to our knowledge, the first demonstration of an

FIG. 4 Transduction of SNO-Hb vasoactivity. **a**, Concentration-effect responses of different SNO-Hb preparations. Contractile effects of Hb(FeII)O₂ (▲) are shown to be partially reversed by S-nitrosylation (SNO-Hb(FeII)O₂ (■); $P = 0.02$ by ANOVA versus Hb(FeII)O₂). Oxidation of the metal centre of SNO-Hb (SNO-Hb(FeIII) (●)) converts the protein into a vasodilator ($P < 0.0001$ by ANOVA versus SNO-Hb(FeII)O₂), with potency comparable to that of other S-nitrosothiols⁴. The contractile properties of Hb(FeII) are shown for comparison (□); $n = 6–17$ for each data point. **b**, Potentiation of SNO-Hb effects by glutathione. Addition of glutathione (100 μ M) to bioassay chambers potentiates the dose-response to both SNO-Hb(FeII)O₂ (■) and SNO-Hb(FeIII) (●) ($n = 6–12$; $P < 0.0001$ for both by ANOVA, compared with the respective tracings in **a**). Glutathione has a transient effect on baseline tone in some experiments, but does not significantly influence the response to Hb(FeII)O₂ (▲). **c**, Transnitrosation between SNO-Hb and glutathione. Rates of NO group transfer from SNO-Hb (100 μ M) to glutathione (10 mM) are displayed for SNO-Hb(FeII)O₂ (oxy) and SNO-Hb(FeIII) (met) ($n = 5$). Data are presented as the amount of GSNO formed relative to the starting SNO-Hb concentration. The transfer is faster for SNO-Hb(FeIII) than SNO-Hb(FeII)O₂ ($P < 0.002$ by ANOVA), suggesting that the GSNO/SNO-Hb equilibrium is shifted towards GSNO by formation of methb. **d**, Export of S-nitrosothiols by red blood cells. Human red blood cells containing SNO-Hb export low-molecular-weight RSNOs over time. Haemolysis, which ranged from 0–< 0.5% over 1 h and did not correlate with rates of RSN release, could account for only a trivial fraction of the measured extracellular RSN.

METHODS. **a**, Details of the vessel ring bioassay have been published elsewhere⁴. SNO-Hb(FeII)O₂ preparations were synthesized with 10-fold excess S-nitrosocysteine over protein (2% borate, 0.5 mM EDTA; ~15 min incubation), then desalted on Sephadex G-25 columns. CYSNO was synthesized in 0.5 N HCl, 0.5 mM EDTA and then neutralized (1:1) in 1 M phosphate buffer containing 0.5 mM EDTA. SNO-Hb(FeIII) preparations followed a similar protocol, but used Hb(FeIII) as starting material. The latter was synthesized by treatment of Hb(FeII)O₂ with excess ferricyanide,



then desalted on G-25 columns. SNO-Hb concentrations were verified spectroscopically and the S-nitrosothiol content determined⁴. The S-NO/tetramer stoichiometry for both SNO-Hb preparations was ~2. **c**, Thiol/SNO-Hb exchange, forming GSNO, was verified chemically⁴ following TCA precipitation ($n = 5$). These results were verified in separate experiments by measuring the residual SNO-Hb concentration after separation of reaction mixtures on G-25 columns. **d**, Packed human red blood cells were washed and resuspended in PBS containing 5 mM SNO-CYS (0.5 mM EDTA, pH 7.4) for 1 h. This results in a red cell preparation containing SNO-Hb(FeII)O₂/FeIII mixture) with a stoichiometry of ~0.5 S-NO/tetramer. The red blood cells were then washed repeatedly to remove residual CYSNO (verified), and incubated in Krebs' solution (1:4). Accumulated extracellular RSN was measured over time⁴. Haemolysis was determined by spectral analysis of red blood cell supernatants following centrifugation.

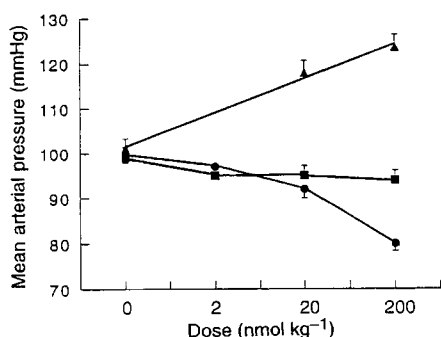


FIG. 5 SNO-Hb bioactivity *in vivo*. *In vivo* effects of cell-free Hb and SNO-Hbs. Administration of 2–200 nmol per kg Hb(FeII)O₂ (▲) (as a bolus) into the femoral vein of a Sprague-Dawley rat increases mean arterial pressure in a dose-dependent manner. At 200 nmol kg⁻¹, mean arterial pressure increased by 20 ± 3 mm Hg (*n* = 4; *P* < 0.05). Elevations in blood pressure reversed within 10–15 min. SNO-Hb(FeII)O₂ infusions (■) (over the same dose range) did not change blood pressure; that is, S-nitrosylation of Hb(Fe II) O₂ ameliorated its hypertensive effects. A similar response was seen at higher doses. By comparison, SNO-Hb(FeIII) infusions (●) caused a significant fall in mean arterial pressure (pre-: 108 ± 4 mm Hg; post-: 74 ± 6 Hg, *n* = 5; *P* < 0.05) at the highest dose (200 nmol kg⁻¹). Hypotensive responses tended to be transient, with blood pressure normalizing over 10 min. A fall in blood pressure was also seen with injection of erythrocytes containing SNO-Hb (see text).

METHODS. Rats were anaesthetized by intraperitoneal injection of pentobarbital and the femoral arteries and veins accessed by local cut down. The artery was then cannulated and the blood pressure monitored continuously using a Viggo Spectramed pressure transducer attached to a Gould recorder. An IBM PC (DATA Q Codas) was used for data acquisition.

intracellular S-nitrosoprotein and draws attention to their role in cellular regulation. The question arises as to how SNO-Hb relaxes blood vessels when any free NO released would be scavenged instantaneously by Hb itself¹. In this regard RSNO activity involves nitrosyl (NO⁺) transfer to thiol acceptors^{9,16,19}, which protect the NO group from inactivation at metal centres. Our

findings indicate that S-nitrosothiol/thiol exchange with glutathione (forming GSNO) occurs within erythrocytes, and is influenced by the oxidation state of haem and its occupation by ligand. Certain activities of GSNO in bacteria require transport of intact dipeptide (that is, S-nitrosocysteinylglycine) across the cell membrane⁵. Our data show that S-nitrosothiol transport happens also in eukaryotic cells. GSNO, or related thiol carriers exported by erythrocytes²⁰, might also initiate signalling in or at the plasmalemma⁶, given reports of thiol-dependent activation of potassium channels by EDRF²¹. An alternative possibility is that Hb associates with red cell membranes through Cysβ93 (ref. 22), which would place Hb in a position to donate the NO group directly to contacting endothelial surfaces, perhaps by SNO/SH exchange. Indeed, cell-surface interactions appear to be operative in signalling mediated by other S-nitrosoproteins^{4,6}.

The highly conserved Cysβ93 residues in Hb influence the oxygen affinity and redox potential of the haem iron and its physicochemical properties^{12–14,23}, but their long sought-after physiological function has remained a mystery. Our studies suggest new sensory and regulatory roles for Hb, in which Cysβ93 functions in transducing NO-related signals to the vessel wall. In particular, the physiological function of Cysβ93, which is invariant in all mammals and birds, is to deliver under allosteric control, NO-related biological activity that cannot be scavenged by haem. Thus, these data bring to light a dynamic circuit for the NO group in which intraerythrocytic Hb acts as both a sink and a donor, depending on its microenvironment. This may explain the paradoxes that arise from concepts based solely on diffusional spread and reaction of free NO (refs 1,2) and may have implications for other thiol- and metal-containing (haem) proteins, such as nitric oxide synthase and guanylate cyclase.

Our discoveries may have therapeutic value. Specifically, concerns over loss of NO-related activity due to inactivation by blood Hb¹ are relieved by the presence of an RSNO subject to allosteric control. Indeed, SNO-Hb is free of the adverse hypertensive properties of cell-free Hb preparations that result from NO scavenging at the haem iron. We suggest, for example, that a cell-free SNO-Hb-containing solution that mimicks blood may be useful as a blood substitute. □

Received 6 November 1995; accepted 14 February 1996.

1. Lancaster, J. R. *Proc. natn. Acad. Sci. U.S.A.* **91**, 8137–8141 (1994).
2. Wood, J. & Garthwaite, J. *Neuropharmacology* **33**, 1235–1244 (1994).
3. Gaston, B. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 10957–10961 (1993).
4. Stamler, J. S. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 444–448 (1992).
5. DeGroot, M. A. et al. *Proc. natn. Acad. Sci. U.S.A.* **92**, 6399–6403 (1995).
6. Stamler, J. S. *Cell* **78**, 931–936 (1994).
7. Olson, J. S. *Meth. Enzym.* **76**, 631–651 (1981).
8. Antonini, E. & Brunori, M. in *Hemoglobin and Myoglobin in their Reactions with Ligands* 29–31 (Elsevier, New York, 1971).
9. Scharfstein, J. S. et al. *J. clin. Invest.* **94**, 1432–1439 (1995).
10. Clancy, R. M. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 3680–3684 (1994).
11. Toothill, C. *Brit. J. Anaesth.* **39**, 405–412 (1967).
12. Garel, C. et al. *J. Biochem.* **123**, 513–519 (1982).
13. Jocelyn, P. C. *Biochemistry of the SH Group* 243 (Academic, London, 1972).
14. Craescu, C. T. et al. *J. biol. Chem.* **261**, 14710–14716 (1986).
15. Wennmalm, A., Benthin, G. & Petersson, A. S. *Br. J. Pharmac.* **106**(3), 507–508 (1992).

16. Arnette, D. R. & Stamler, J. S. *Arch. Biochem. Biophys.* **318**, 279–285 (1995).
17. Vogel, W. M. et al. *Am. J. Physiol.* **251**, H413–H420 (1986).
18. Olsen, S. B. et al. *Circulation* **93**, 329–332 (1996).
19. Stamler, J. S. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 7674–7677 (1992).
20. Kondo, T., Dale, G. L. & Beutler, E. *Meth. Enzym.* **252**, 72–83 (1995).
21. Bolotina, V. M. et al. *Nature* **368**, 850–853 (1994).
22. Salhany, J. M. & Gaines, K. C. *Trends biochem. Sci.* Jan, 1–3 (1981).
23. Mansouri, A. *Biochem. biophys. Res. Commun.* **89**, 441–447 (1979).
24. Kilbourn, R. G. et al. *Biochem. biophys. Res. Commun.* **199**, 155–162 (1994).
25. Beckman, J. S., Wink, D. & Crow, J. P. in *Methods in Nitric Oxide Research* (eds Feelisch, M. & Stamler, J. S.) (Wiley, Chichester, UK, 1996).
26. Shen, W., Hintze, T. H. & Wolin, M. S. *Circulation* **92**, 3505–3512 (1995).

ACKNOWLEDGEMENTS. We thank M. Dewhurst and I. Fridovich for use of equipment, and J. DeAngelo for discussion. J.S.S. is a Pew Scholar in the Biomedical Sciences and the recipient of a Clinical Investigator Award from the NIH. This work was also funded by a grant from the NIH and a centre grant from the NIEHS.

Infrared emission spectra of candidate interstellar aromatic molecules

D. J. Cook, S. Schlemmer, N. Balucani, D. R. Wagner, B. Steiner & R. J. Saykally*

Department of Chemistry, University of California, Berkeley, California 94720-1460, USA

*To whom correspondence should be addressed.

INTERSTELLAR dust is responsible, through surface reactions, for the creation of molecular hydrogen, the main component of the interstellar clouds in which new stars form. Intermediate between small, gas-phase molecules and dust are the polycyclic aromatic hydrocarbons (PAHs). Such molecules could account for 2–30% of the carbon in the Galaxy¹, and may provide nucleation sites for the formation of carbonaceous dust^{2,3}. Although PAHs have been proposed^{4,5} as the sources of the unidentified infrared emission bands that are observed in the spectra of a variety of interstellar sources^{6–11}, the emission characteristics of such molecules are still poorly understood. Here we report laboratory emission spectra of several representative PAHs, obtained in conditions approximating those of the interstellar medium, and measured over the entire spectral region spanned by the unidentified infrared bands. We find that neutral PAHs of small and moderate size can at best make only a minor contribution to these emission bands. Cations of these molecules, as well as much larger PAHs and their cations, remain viable candidates for the sources of these bands.

The unidentified infrared bands (UIRs) are generally associated with interstellar molecular clouds that are subjected to intense ultraviolet and optical radiation^{6–11}; one such example is the Trapezium region of the Orion Nebula, where the massive young stars (such as θ^1 Orioni C) have ionized the surrounding gas cloud from which they formed. The UIRs are also seen towards planetary nebulae, which result when the gas and dust surrounding a newly formed white dwarf star is ionized and heated by the radiation from the white dwarf.

In the PAH model of the UIRs^{4,5}, absorption of ultraviolet light by an isolated PAH is followed by rapid internal energy conversion, relaxing the molecule to the electronic ground state, but leaving it with the energy of the ultraviolet photon (typically 10 eV) distributed among its vibrational modes. In the collisionless interstellar environment, the vibrationally hot molecule is hypothesized to cool by infrared emission at frequencies characteristic of PAH vibrations, giving rise to the UIRs. UIR frequencies and their PAH local-mode assignments are summarized in Table 1.

We have designed a new experiment to test rigorously the PAH hypothesis in the laboratory. Because of the considerable uncertainties associated with modelling the interstellar environment, our approach is to systematically examine each class of PAH molecule that has been proposed as a possible UIR source. Here we report measurements of infrared emission spectra of four unsubstituted PAHs, two methyl substituted PAHs, and two hetero-substituted PAHs (Fig. 1), all with complete

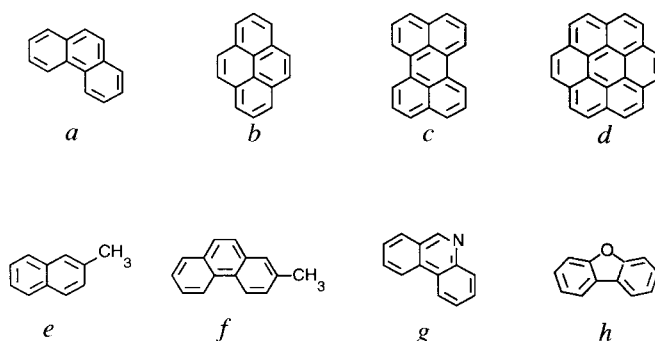


FIG. 1 Chemical structures representing the PAH molecules studied by SPIRES in this work. These include: four unsubstituted PAHs, phenanthrene (a, $C_{14}H_{10}$), pyrene (b, $C_{16}H_{10}$), perylene (c, $C_{20}H_{12}$) and coronene (d, $C_{24}H_{12}$); two methyl substituted PAHs, 2-methylnaphthalene (e, $C_{11}H_{10}$), and 2-methylphenanthrene (f, $C_{15}H_{12}$); and two hetero-substituted PAHs, phenanthridine (g, $C_{13}N_3H_9$) and dibenzofuran (h, $C_{12}H_8O$). Molecules the size of coronene have recently been postulated to account for the 3.3- μ m UIR band²⁰.

coverage of the infrared spectral regions spanned by the UIRs. To obtain such extensive spectral coverage, we use ultraviolet-laser-induced desorption to introduce highly excited gas phase PAHs into a low-infrared-background cryogenic high vacuum chamber. A cryogenic monochromator disperses the infrared emission, and a blocked impurity band solid-state photomultiplier¹² is used to count single infrared photons. Our initial results¹³, in which this single-photon infrared emission spectroscopy (SPIRES) apparatus is described, demonstrate the applicability of this technique to the PAH-UIR problem; monochromator heat leaks limited the spectral coverage in this previous report.

Figure 2 shows SPIRES spectra of the unsubstituted PAHs (Fig. 2a) and substituted PAHs (Fig. 2b) along with the characteristic UIR wavelengths. When comparing laboratory spectra to the UIRs, we note that no single PAH, but rather a family of molecules with similar structure and presumably similar infrared spectra, is postulated to give rise to the UIRs^{14,15}. Bands which appear in a characteristic spectral region but with significant differences in frequency are postulated to combine to produce an apparent continuum, such as the continuum observed in the UIR spectra between 5.5 and 9.5 μ m. But if a conglomerate spectrum is to exhibit relatively narrow bands like the UIR bands at 3.3, 6.2 and 11.2 μ m, then clearly these same narrow

TABLE 1 PAH-UIR hypothesis summary

UIR wavelength (μ m)	3.3	6.2	7.7	8.7	11.2
UIR bandwidth FWHM (cm^{-1})	30	30	70 to 200*		30
Characteristic PAH vibrations (see text)	C–H stretch	C–C stretch	C–C stretch	C–H in-plane bend	C–H out-of-plane bend†

Typical peak wavelengths and bandwidths of the primary UIR bands, along the characteristic PAH local-mode vibrations, from Allamandola et al.⁴, unless stated otherwise. Minor UIR features are not tabulated.

* The 7.7- μ m feature is broad and varies from source to source. The 8.7- μ m band is often seen as a shoulder on this feature.

† The wavelength of the 11.2- μ m UIR band is characteristic of C–H out-of-plane bending motions arising from solo C–H bonds on the periphery of the ring. An *ab initio* survey of PAH ion frequencies²⁷ indicates that for large cations with condensed structures, the 11.2- μ m band may also be characteristic of C–H bonds with one adjacent C–H bond (duo C–H) on the periphery of the ring.

bands must appear consistently in the individual molecular emission spectra. This is not observed in our laboratory spectra for bands arising primarily from C–C stretching and C–H out-of-plane bending vibrations. This constitutes strong evidence that the molecules studied here cannot be major sources of the UIRs.

The 3.3- μm bands observed in our spectra do show a remarkable invariance in frequency among the different PAHs studied, and yield consistent and significant red shifts relative to the bands measured in cryogenic matrix absorption experiments ($\sim 25\text{ cm}^{-1}$). Emission features with similar lineshapes and exhibiting systematic red shifts with increasing internal energy content have been observed in previous ultraviolet laser excitation experiments^{13,16,17}, as well as in high-temperature equilibrium gas cell experiments^{18,20}, and have been attributed to unresolvable sequence bands which are characteristic of highly vibrationally excited molecules^{13,17}. Similar red shifts are observed for the other bands in the spectra, although the effect is not as pronounced. As the red shifts and band broadening due to high internal energy content can only be determined empirically, experimental measurements of molecules under astrophysical conditions are essential to test the PAH-UIR hypothesis. Such measurements are also needed to provide an adequate basis for modelling the

emission spectra which one expects to observe for other molecules, such as PAH ions, for which only matrix data presently exist.

Although relative intensities reported here are only semi-quantitative, gross differences clearly exist between the intensity patterns of the SPIRES spectra and the UIRs. The 6.2- and 7.7- μm UIR bands, assigned primarily to C–C stretches, are strong when compared to 3.3- μm C–H stretches and 11.2- μm C–H out-of-plane bends; our laboratory PAH spectra exhibit the opposite result. The intensity differences reported here indicate that, whereas neutral PAHs such as those included in this study could account for the majority of the UIR emission observed in the 3.3- μm band and the minor 3.4- μm band, they could account for less than 10% of the emission observed in the other UIR bands.

Such intensity differences between neutral PAHs and the UIRs have been noted before¹⁹, and thus constitute a general argument against these molecules being major UIR sources. Recent attention has consequently been turned to ionized PAHs. Matrix absorption spectra of PAH cations^{21–25} produce a promising trend of bands shifting towards the 7.7- μm and 11.2- μm features, as well as band intensity ratios more closely resembling the UIRs. Models predict that PAHs are largely ionized in many regions where the UIRs are generated^{4,26}. Because of the substantial

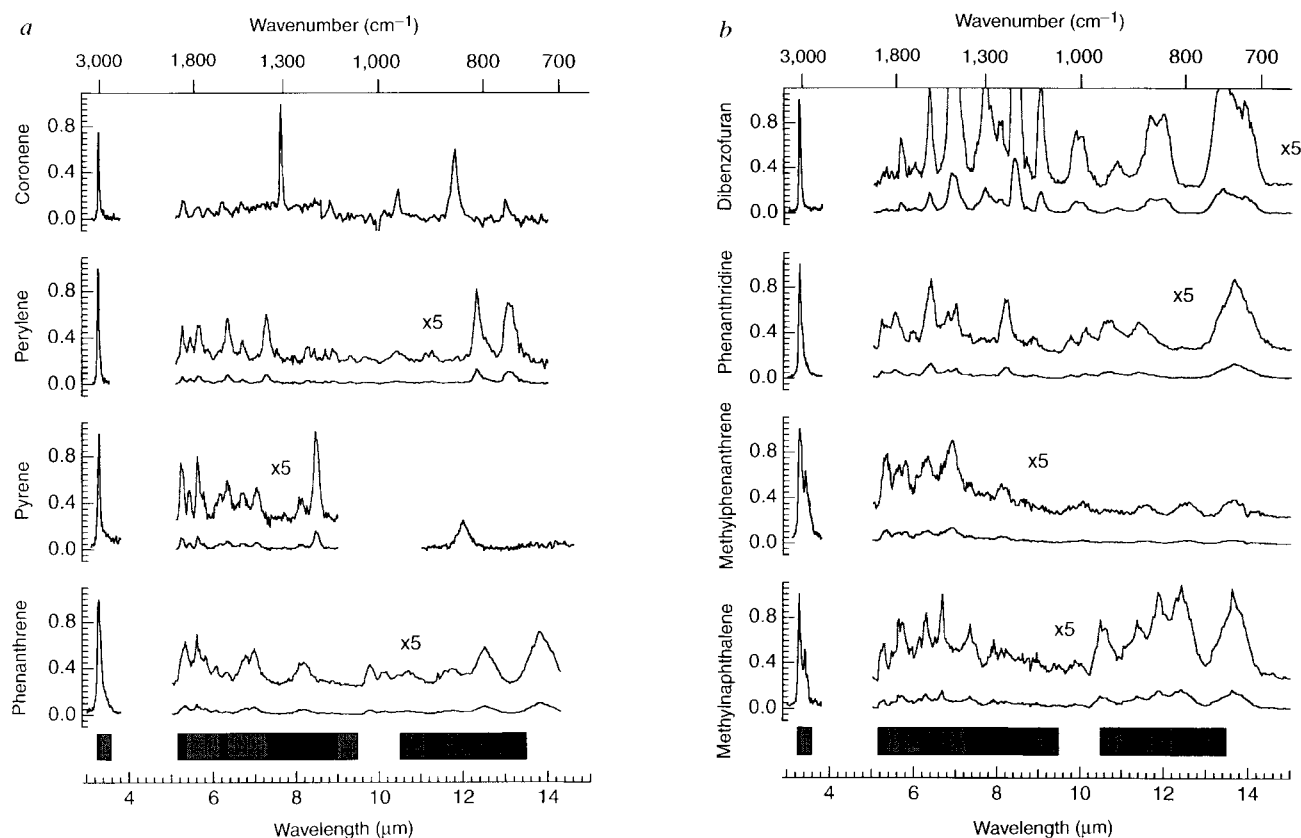


FIG. 2 Infrared emission spectra of the unsubstituted PAHs (a) and the substituted PAHs (b) depicted in Fig. 1, measured by SPIRES under conditions approximating the interstellar environment. The bars at the bottom of the spectra represent the wavelengths at which UIR bands are observed. The dark bands correspond to wavelengths at which distinct peaks are observed, and the light-grey bands correspond to wavelengths where associated continuum emission is observed. For pyrene, perylene and coronene, a 150 groove mm^{-1} grating was used for the spectra between 3 and 4 μm ; between 5 and 14 μm a 75 groove mm^{-1} grating was used. For phenanthrene and the substituted PAHs, a 50 groove mm^{-1} grating was used between 10 and 15 μm . The spectra were corrected for the estimated instrument response functions. Difficulties associated with calibration using a variable-temperature cryogenic black-body source include

limited detector dynamic range and stray light from the ruled gratings, which lead to estimated uncertainties of a factor of ten with regard to relative band intensities. The spectra from the multiple gratings were combined, and the relative intensities were converted from photon flux to power and normalized to the most intense peak in each individual spectrum. A vibrational temperature can be estimated by using the SPIRES frequencies in conjunction with the pyrene and coronene red-shift measurements of Joblin and co-workers¹⁹, which assume a linear dependence of peak position on temperature. Temperatures are calculated for each band and the result is averaged. This calculation yields vibrational temperatures of 1,019 K and 814 K for pyrene and coronene, respectively. In our previous report¹³, we used a similar approach to estimate the internal energy of laser-desorbed naphthalene to be $\sim 30,000\text{ cm}^{-1}$.

differences in ion infrared band strengths with respect to the neutrals²⁷, only a small excess of ions would be necessary to reproduce the UIRs, so long as these ions consistently exhibit strong features at 6.2 and 11.2 μm , and exhibit a lack of strong features at other non-UIR wavelengths. Hence it appears essential to extend the laboratory experiments described here to ionized PAHs to assess definitively the PAH hypothesis. We note that the general mechanism proposed for production of the UIRs (ultraviolet absorption followed by internal conversion and infrared emission)^{4,5} is quite strongly supported by our measurements, as well as by earlier measurements of ultraviolet-laser-excited PAHs^{13,16,17,28,29} in the 3- μm region. $\square$

Received 20 June 1995; accepted 6 February 1996.

1. Snow, T. P. & Witt, A. N. *Science* **270**, 1455–1460 (1995).
2. Tielens, A. G. G. M. in *Carbon in the Galaxy: Studies From Earth and Space* (eds Tarter, J. C., Chan, S. & DeFrees, D. J.) 59–111 (NASA Conf. Publ. 3061, Moffett Field, 1987).
3. Cadwell, B. J., Wang, H., Feigelson, E. D. & Frenklach, M. *Astrophys. J.* **429**, 285–299 (1994).
4. Allamandola, L. J., Tielens, A. G. G. M. & Barker, J. R. *Astrophys. J. Suppl. Ser.* **71**, 733–775 (1989).
5. Léger, A. & d'Hendecourt, L. *Ann. Phys. Fr.* **14**, 181–206 (1989).
6. Cohen, M. et al. *Astrophys. J.* **341**, 246–269 (1989).
7. Geballe, T. R., Lacy, J. H., Persson, S. E., McGregor, P. J. & Soifer, B. T. *Astrophys. J.* **292**, 500–505 (1985).

8. Witteborn, F. C. et al. *Astrophys. J.* **341**, 270–277 (1989).
9. de Muizon, M. J., d'Hendecourt, L. B. & Geballe, T. R. *Astr. Astrophys.* **235**, 367–378 (1990).
10. Schutte, W. A., Tielens, A. G. G. M., Allamandola, L. J., Cohen, M. & Wooden, D. H. *Astrophys. J.* **360**, 577–589 (1990).
11. Allamandola, L. J. et al. *Astrophys. J.* **345**, L59–L62 (1989).
12. Petroff, M. D., Stapelbroek, M. G. & Kleinmans, W. A. *Appl. Phys. Lett.* **51**, 406–408 (1987).
13. Schlemmer, S. et al. *Science* **265**, 1686–1689 (1994).
14. Allamandola, L. J., Tielens, A. G. G. M. & Barker, J. R. *Astrophys. J.* **290**, L25–L28 (1985).
15. Léger, A. & Puget, J. L. *Astr. Astrophys.* **137**, L5–L8 (1984).
16. Shan, J., Suto, M. & Lee, L. C. *Astrophys. J.* **383**, 459–465 (1991).
17. Williams, R. M. & Leone, S. R. *Astrophys. J.* **443**, 675–681 (1995).
18. Flickinger, G. C., Wdowiak, T. J. & Gómez, P. L. *Astrophys. J.* **380**, L43–L46 (1991).
19. Joblin, C., d'Hendecourt, L., Léger, A. & Défourneau, D. *Astr. Astrophys.* **281**, 923–936 (1994).
20. Joblin, C., Boissel, P., Léger, A., d'Hendecourt, L. & Défourneau, D. *Astr. Astrophys.* **299**, 835–846 (1995).
21. Allamandola, L. J., Sandford, S. A., Hudgins, D. M. & Witteborn, F. C. in *Airborne Astronomy Symposium on the Galactic Ecosystem* (eds Haas, M. R., Davidson, J. A. & Erickson, E. F.) 23–32 (Astronomical Soc. of the Pacific, San Francisco, 1995).
22. Hudgins, D. M. & Allamandola, L. J. *J. phys. Chem.* **99**, 3033–3046 (1995).
23. Szczepanski, J. & Vala, M. *Astrophys. J.* **414**, 646–655 (1993).
24. Szczepanski, J., Chappo, C. & Vala, M. *Chem. Phys. Lett.* **205**, 434–439 (1993).
25. Vala, M. et al. *J. phys. Chem.* **98**, 9187–9196 (1994).
26. Omont, A. *Astr. Astrophys.* **164**, 159–178 (1986).
27. Langhoff, S. R. *J. phys. Chem.* (in the press).
28. Brenner, J. & Barker, J. R. *Astrophys. J.* **388**, L39–L43 (1992).
29. Cherchneff, I. & Barker, J. R. *Astrophys. J.* **341**, L21–L24 (1989).

ACKNOWLEDGEMENTS. This work was supported by the NASA Astrophysics Program, the NASA Exobiology Program, and fellowships from the Deutsche Forschungsgemeinschaft (S.S.), the Consiglio Nazionale delle Ricerche (Italy) (N.B.) and the Max-Planck-Gesellschaft and the Alexander von Humboldt Stiftung (B.S.).

Discovery of an extended sodium atmosphere around Europa

Michael E. Brown* & Richard E. Hill

Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

EUROPA, one of the satellites of Jupiter, has long been thought to be a dormant icy body¹, unlike its volcanically active neighbour, Io. Europa lies deep within Jupiter's magnetosphere, however, and is continuously bombarded by energetic ions, which modify the surface ices² and are probably responsible for creating Europa's tightly bound oxygen atmosphere^{3,4}. Here we report the discovery of an atmosphere of atomic sodium that extends to at least 25 times Europa's radius. We suggest that this sodium is originally released by Io's volcanoes, after which it is ionized in the magnetosphere and implanted into Europa's surface ice; subsequent sputtering of the ice by magnetospheric ions releases the sodium to form the extended atmosphere. Although sodium is a minor constituent of Europa's atmosphere, it traces the distribution of the major atmospheric components which are not themselves directly observable. The sodium and oxygen could represent the extremes of the distribution of the atmospheric components, with only the heaviest molecules (such as the oxygen) being tightly bound; alternatively the sodium might be in the form of an extended corona, analogous to Io's atmosphere.

Observations of sodium around Europa were made on 5 June 1995 using an echelle spectrograph⁵ on the 1.53-m University of Arizona telescope on Mt. Bigelow. We obtained ten long-slit high-resolution (0.25 Å) spectra of Europa, each with the slit oriented perpendicular to Europa's orbital plane, and covering the wavelengths between 5,883 and 5,904 Å, which includes the sodium D₁ and D₂ lines at 5,895.92 and 5,889.95 Å, respectively. Table 1 gives a log of the observations. All spectra were obtained when Europa was far from Io to minimize contamination from sodium emitted from Io, but faint Io sodium is still likely to be present as a

spectrally and spatially diffuse background. Observations of the disk of Jupiter were obtained for intensity calibration. (We assume 5.5 MR Å⁻¹ for the peak intensity of the disk⁶, where 1 R = (1/4 π) $\times$ 10⁶ photons cm⁻² s⁻¹ sr⁻¹.) Because terrestrial sodium emission (intensity $\sim$ 100 R) is uniform across the spatial dimension of the slit and is Doppler-shifted from Europa sodium by -0.15 Å (-7 km s⁻¹), it is readily subtracted. The same subtraction removes the uniformly distributed Io sodium. Figure 1 shows a typical spectrum of Europa and of the sodium emission. The observed emission lines are centred at wavelengths of 5,895.77 and 5,889.80 Å, precisely those expected for sodium at Europa's velocity, and the emission is strongly peaked at the spatial position of Europa. We therefore conclude that Europa is the source of the sodium. No changes are observable during the

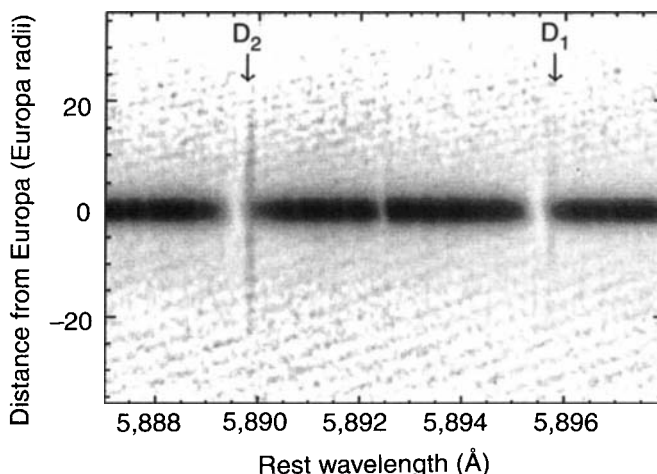


FIG. 1 A long-slit high-resolution spectrum of Europa showing sodium emission. The dark horizontal band is a spectrum of reflected sunlight from Europa's surface in which solar Fraunhofer absorption features are visible (including the strong sodium absorptions). The sodium emission lines are marked by D₂ and D₁ and can be traced to at least 25 Europa radii. The spectral slit was oriented perpendicular to the orbital plane, so the sodium emission seen is from north and south of the satellite. This spectrum has been heavily processed to show the faint sodium emission against the much brighter Europa continuum, and a constant contribution from atmospheric sodium emission has been subtracted.

* Present address: Division of Geological and Planetary Sciences, Mail Stop 170-25, California Institute of Technology, Pasadena, California 91125, USA.

4.5 hours of observations, so we add together all spectra to increase the signal-to-noise ratio.

The observed spatial distribution of sodium emission around Europa is shown in Fig. 2. Because of the overwhelming brightness of the Europa continuum, the emission intensity inside $\sim 5 R_e$ (where R_e is a distance equal to Europa's 1,569-km radius) cannot be quantified. The ratio of the intensities of the D_2 and D_1 lines is 1.70 ± 0.05 , consistent with the expected value of 1.66 for optically thin emission and resonant scattering of sunlight⁶; therefore the line-of-sight column density of sodium (shown on the right-hand vertical axis of Fig. 2) is directly proportional to the emission intensity. Figure 3 shows the sodium column density on a log-log scale to illustrate better the atmospheric structure. The best-fit line through these points has a slope of -1.5 ± 0.1 . If the sodium is distributed symmetrically around Europa, this slope implies a total mass of sodium between 1 and $25 R_e$ of ~ 840 kg and a density at Europa's surface of ~ 70 atoms cm^{-3} . This sodium surface density is 100 times smaller than the extrapolated surface density of sodium at Io⁷.

The extent of the sodium distribution implies that atoms are leaving the surface at high velocity: to reach a height of $25 R_e$ above a pole of Europa requires a velocity greater than 2 km s^{-1} . Preliminary modelling of particle trajectories suggests that most of the particle velocities must be this large or larger to avoid the accumulation of gravitationally bound particles close to Europa and account for the flatness of the profile between 5 and $10 R_e$. As sublimation at the Europa surface temperature of 95 K will produce typical velocities of only 0.3 km s^{-1} , the most likely mechanism for generating the required high velocities is sputtering by energetic magnetospheric particles⁴.

Sodium must be present on Europa's surface in order to be sputtered into the atmosphere. A potential sodium source is ions transported outward from Io and implanted into the surface ice of Europa. Knowing the ion implantation rate and the lifetime for atmospheric sodium, we can calculate the total mass due to this source and compare it to the observed sodium mass. At Europa, the magnetospheric ion density is ~ 80 ions cm^{-3} (ref. 8), of which

TABLE 1 Observations of Europa spectra

Start time (μr)	Exposure time (min)	Air mass	Europa phase (deg)	Io phase (deg)
04:11	20	2.60	91	227
04:34	20	2.21	93	230
05:24	20	1.91	96	237
05:47	30	1.81	98	240
06:18	30	1.72	100	245
06:58	30	1.68	103	251
07:30	30	1.70	105	255
08:09	30	1.78	108	261
08:42	30	1.90	110	265
09:21	20	2.18	113	271

$\sim 1\%$ is thought to be sodium⁷, and the ions sweep past at $\sim 100 \text{ km s}^{-1}$ (ref. 9), producing an implantation rate into Europa's surface of 24 g s^{-1} of sodium. All of this sodium is subsequently removed by sputtering: implantation depths are at most $\sim 100 \text{ \AA}$, so the sodium is removed within at most 100 yr (ref. 4). Once removed from the surface, neutral sodium will be destroyed by electron impact ionization or transported beyond the $25 R_e$ radius that we have observed. The lifetime of neutral sodium against electron impact ionization is between 20 and 100 h, depending on the local magnetospheric electron density^{8,10}. The transport time for particles with initial velocities between 2 and 2.4 km s^{-1} to escape past $25 R_e$ is 10–30 h. Assuming a mean combined lifetime of 20 h due to both processes, the steady-state mass of sodium surrounding Europa from this source should be $\sim 1,700$ kg, within a factor of two of the observed amount.

Naturally occurring sodium impurities in Europa's surface ice might also contribute to the atmosphere. Assuming the cosmic Na:O ratio of 1:370 (ref. 11) and using Johnson's estimate for the water sputtering rate⁴ of 10^8 – 10^{10} molecules $\text{cm}^{-2} \text{ s}^{-1}$, the sodium removal rate is 1–100 g s^{-1} . At the upper limit of the range, this source exceeds the implantation source, but the sputtering rate and intrinsic sodium abundance are both highly uncertain. Because of this uncertainty and because the implantation source must be present at approximately the levels calculated, we conclude that implantation and sputtering of Io sodium is probably the dominant source of the Europa sodium. If true, this sodium atmosphere is the first known example of an atmospheric constituent captured from an external moon or planet.

Although sodium is only a minor species, with a total mass 300

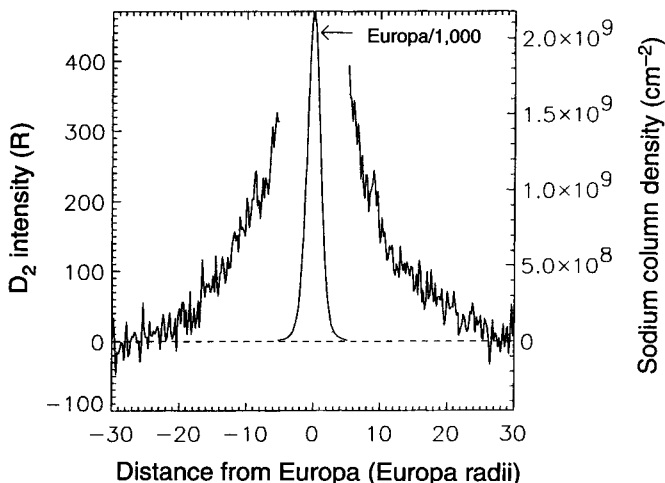


FIG. 2 A profile of the D_2 sodium emission intensity and of the sodium column density. The smooth curve labelled 'Europa/1,000' shows the emission intensity of Europa, divided by 1,000, and demonstrates the spatial resolution of the data. Inside 5 Europa radii, the sodium emission intensity cannot be determined because it is overwhelmed by the much brighter Europa continuum.

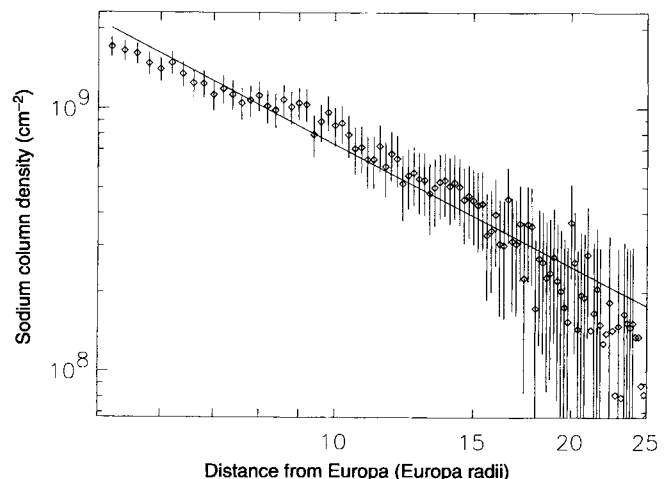


FIG. 3 A log-log plot of the sodium column density. The best-fit line through these data has a slope of -1.5 . Assuming spherical symmetry, the total mass implied by this slope is 840 kg.

times less than that of the closely confined O_2 atmosphere, the sodium traces the behaviour of the dominant sputtered water products that are not directly observable. These observations therefore show that the Europa atmosphere contains a significant escaping component in addition to the bound component previously observed. These two components could represent the extremes of a single energy distribution, with the more massive O_2 having low velocity and remaining bound, and the lighter sodium atoms (and even lighter water products) acquiring high velocity and escaping. In this case, the atmosphere would be predominantly escaping with only the heavy O_2 tightly bound. Alternatively, Europa's atmosphere may be more analogous to Io's, with a bound collisionally thick lower atmosphere and an extended corona sputtered from this bound component. In this case, the bound atmosphere would contain significant amounts of water products and of sodium, in addition to the observed O_2 , and the sputtered corona would contain a representative sample of this atmosphere. Further understanding of the structure and nature of Europa's atmosphere will require observations to link the bound and escaping components, detecting either oxygen further out or sodium closer in to the satellite. The accessibility of sodium emissions to ground-based telescopes suggests that sodium observations will be most useful for advancing our understanding of this satellite and its atmosphere. $\square$

Received 28 July 1995; accepted 29 January 1996.

1. Clark, R. N., Fanale, F. P. & Gaffey, M. J. in *Satellites* (eds Burns, J. A. & Matthews, M. S.) 437–491 (Univ. Arizona Press, Tucson, 1986).
2. Johnson, R. E., Nelson, M. L., McCord, T. B. & Gradie, J. C. *Icarus* **75**, 423–436 (1988).
3. Hall, D. T., Strobel, D. F., Feldman, P. D., McGrath, M. A. & Weaver, H. A. *Nature* **373**, 677–679 (1995).
4. Johnson, R. E. *Energetic Charged-Particle Interactions with Atmospheres and Surfaces* (Springer, Berlin, 1990).
5. Hunten, D. M., Wells, W. K., Brown, R. A., Schneider, N. M. & Hilliard, R. L. *Publ. astr. Soc. Pacif.* **103**, 1187–1192 (1991).
6. Brown, R. A. & Yung, Y. L. in *Jupiter* (ed. Gehrels, T.) 1102–1145 (Univ. Arizona Press, Tucson, 1976).
7. Schneider, N. M., Hunten, D. M., Wells, W. K., Schultz, A. B. & Fink, U. *Astrophys. J.* **368**, 298–314 (1991).
8. Bagenal, F. J. *geophys. Res.* **99**, 11043–11062 (1994).
9. McNutt, R. L., Belcher, J. W., Sullivan, J. D., Bagenal, F. & Bridge, H. S. *Nature* **280**, 803–804 (1979).
10. Smyth, W. H. & Combi, M. *Astrophys. J.* **328**, 888–918 (1988).
11. Allen, C. W. *Astrophysical Quantities* (Athlone, London, 1973).

ACKNOWLEDGEMENTS. We thank E. J. Moyer, A. J. Dessler and D. M. Hunten for enlightening conversations. This work was supported by a Hubble Fellowship grant from STScI.

Yttrium and lanthanum hydride films with switchable optical properties

J. N. Huiberts, R. Griessen, J. H. Rector, R. J. Wijngaarden, J. P. Dekker, D. G. de Groot & N. J. Koeman

Faculty of Physics and Astronomy, Vrije universiteit, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands

In many substances, changes in chemical composition, pressure or temperature can induce metal-to-insulator transitions¹. Although dramatic changes in optical and electrical properties accompany such transitions, their interpretation is often complicated by attendant changes in crystallographic structure². Yttrium, lanthanum and the trivalent rare-earth elements form hydrides that also exhibit metal–insulator transitions^{3–5}, but the extreme reactivity and fragility of these materials hinder experimental studies^{5,6}. To overcome these difficulties, we have coated thin films of yttrium and lanthanum with a layer of palladium through which hydrogen can diffuse. Real-time transitions from

metallic (YH_2 or LaH_2) to semiconducting (YH_3 or LaH_3) behaviour occur in these films during continuous absorption of hydrogen, accompanied by pronounced changes in their optical properties. Although the timescale on which this transition occurs is at present rather slow (a few seconds), there appears to be considerable scope for improvement through the choice of rare-earth element and by adopting electrochemical means for driving the transition. In view of the spectacular changes in optical properties—yttrium hydride, for example, changes from a shiny mirror to a yellow, transparent window—metal hydrides might find important technological applications.

In some cases metal–insulator transitions can be induced by varying continuously an externally controlled parameter such as pressure or temperature. A well known example is that of VO_2 which exhibits a semiconductor-to-metal transition around 340 K during heating⁷. At the transition the electrical conductivity changes abruptly although the optical properties at visible wavelengths are only weakly modified. From a fundamental point of view, the interpretation of the metal–insulator transition in VO_x is complicated by the drastic crystallographic changes (distorted monoclinic to tetragonal) which occur at the transition. Another class of solids exhibiting a metal–insulator transition is that of the trihydride-forming metals, Y, La and the trivalent rare-earths (RE)^{3–5}.

Unfortunately bulk samples of YH_x , LaH_x and REH_x often fall apart into powder when x approaches 3 (refs 5,6), severely limiting the possibility of investigating their electrical and optical properties. The nature of the electronic ground state of the trihydrides has accordingly not been firmly established⁷. By evaporating films one might improve the structural stability of the trihydrides. However, thin films of Y, La and RE are extremely reactive and thus difficult to handle. To circumvent these difficulties, we evaporated yttrium films under ultra-high vacuum (UHV; 10^{-6} Pa during evaporation) conditions and coated them with a thin layer of palladium (5–20 nm thick) through which hydrogen may pass to, or from, the (typically 500-nm-thick) yttrium film. The palladium overlayer protects the yttrium film from oxidation so effectively that the samples could be transferred without special precautions to *ex situ* experimental setups. These samples made possible a comprehensive study of properties of YH_x ($0 \leq x < 3$), such as electrical resistivity ρ , magnetoresistance, photoconductivity, Hall effect and optical transmission as a function of temperature (between 4.2 and 300 K), magnetic field (between 0 and 7 T) and composition (J.N.H. *et al.*, manuscript in preparation). As far as a comparison with bulk YH_x is possible (for electrical resistivity data, x is limited to < 2.1) our YH_x films exhibit the characteristic features of bulk YH_x samples. For example, the temperature dependence of ρ exhibits the 175 K anomaly in the α^* - YH_x phase⁸ and the hydrogen-concentration dependence of ρ shows a pronounced minimum in the β - YH_2 phase and a rapid increase in the γ - YH_3 phase. The lattice spacings of the film (a) are also very close to the bulk values (for example, $a = 5.219$ Å (YH_2 -film), $a = 5.205$ Å (ref. 9), 5.201 Å (ref. 10), 5.209 Å (ref. 11) (bulk YH_2)). Thermodynamically, YH_x films are slightly less stable than bulk YH_x , the enthalpy of formation for our films $\Delta H_{\beta-\gamma} \approx -38$ kJ per mol H being 6 kJ per mole H higher than that of bulk samples¹². This well known effect is due to the fact that the effective attractive H–H interaction in metal hydrides is essentially of elastic origin¹³. The partial clamping of the film by the substrate leads to a weakening of the H–H interaction and consequently to an increase in $\Delta H_{\beta-\gamma}$. For the present study, this is an advantage because it allows the removal of hydrogen from YH_3 to give YH_2 without having to use UHV pumping. With our thin-film technique we have also recently succeeded in producing stoichiometric YH_3 by loading Y under very high pressures (up to 25 GPa of hydrogen) in a diamond anvil cell. The changes in properties of YH_x as a function of x obtained by simply subjecting a 500-nm-thick film of Y (with a 20-nm Pd cap layer) to 10^5 Pa of hydrogen gas are, however, very interesting, and we describe them here.

A vivid demonstration of the optical changes that arise upon hydrogen loading at room temperature under 10^5 Pa H_2 is given in Fig. 1a–c. At room temperature, within a few minutes the perfectly reflecting pure yttrium film (Fig. 1a) transforms into a partially reflecting dihydride (Fig. 1b) and later into an essentially non-reflecting and transparent yellowish $YH_{2.86}$ film. The concentration $x = 2.86 \pm 0.02$ has been determined independently from a quartz crystal microbalance^{14,15} and nuclear ^{15}N profiling¹⁶ measurements. The change of contrast during the transition is much larger than that of (for example) VO_2 in the visible range. The transition is also completely different from that of hydrogen intercalated bronzes¹⁷ (for example, H_xWO_3) which change from transparent (colourless) to transparent deep blue without exhibiting a true metal–insulator transition.

To investigate the properties of the YH_x film during continuous hydrogen uptake for $0 < x < 2.86$, we simultaneously carried out electrical resistivity and light transmission measurements ($1\text{ eV} \leq \hbar\omega \leq 3.2\text{ eV}$). As an example, Fig. 2b shows the light transmission through the sample at one photon energy ($\hbar\omega = 1.8\text{ eV}$) during the hydrogen absorption. The measurements of these properties are shown as a function of time. At $t = 0$, the hydrogen gas pressure is changed from 0 to 0.9×10^5 Pa. As the hydrogen concentration in Y is increasing with time, the curves in Fig. 2a and b can be qualitatively viewed as curves of electrical resistivity and optical transmission, respectively, versus hydrogen concentration (Fig. 2c). Immediately after $t = 0$, resistivity increases owing to (1) a decrease in the number of free charge carriers and (2) additional electron scattering due to randomly distributed hydrogen. At $t = 17$ s, dihydride (β) phase starts to precipitate, and the resistivity decreases as the dihydride

has an approximately 5 times lower resistivity than pure yttrium. The material undergoes a structural phase transformation from the α^* (hexagonal close packed (h.c.p.) yttrium) to the β (face-centred cubic (f.c.c.) yttrium dihydride) phase. The various phase transitions of the metal sublattice during hydrogen loading in our films were confirmed by X-ray measurements (J.N.H. *et al.*, manuscript in preparation). At higher concentrations ($t = 65$ s) the resistivity increases rapidly, first due to an increase of octahedral occupation in the f.c.c. (β - YH_2 phase) dihydride followed by a further increase due to the formation of h.c.p. (γ - YH_3 phase) trihydride. The difference in resistivity between YH_2 and $YH_{2.86}$ slightly exceeds two orders of magnitude, clearly suggesting that $YH_{3-\delta}$ is semiconducting (J.N.H., manuscript in preparation).

Further evidence for this is the optically measured absorption edge, from which we conclude that $YH_{3-\delta}$ has a gap of 1.8 eV with an edge extending up to 2.8 eV for films of 500 nm thickness. The absorption coefficient α rises from $\sim 1\text{ cm}^{-1}$ for $\hbar\omega \leq 1.8\text{ eV}$ ($\sim 100\%$ transmission) to 10^5 cm^{-1} at 2.8 eV (0.1% transmission). The position of the edge explains the yellow colour of $YH_{2.86}$ seen in transmission in Fig. 1c.

The optical transmission data in Fig. 2b at $t = 65$ s exhibit a remarkably fast (few seconds) increase in transmission close to $x \approx 2$. The film is moderately transparent ($\sim 1\%$ transmission after correction for the Pd layer) in a small interval around the resistivity minimum. At slightly higher concentrations, owing to an extra absorption related to octahedral site occupancy^{18,19}, transmission drops again to zero. Further increase of the hydrogen concentration leads to an abrupt and drastic increase in optical transmission intensity at $t = 80$ s. Comparison of our resistivity



FIG. 1 These photographs show the behaviour of a 500-nm-thick yttrium film covered with a 20-nm palladium protection layer. The film and the transparent chess-board pattern behind it are illuminated from behind to monitor the frequency dependence of the transmitted light, and from in front to see the mirror image of the white knight that is placed just in front of the film. The illumination is the same for all three photographs. The yttrium film, the chess-board pattern and the knight are placed within a glass tube of 20 cm diameter which can be evacuated and filled with hydrogen gas at pressures up to 10^5 Pa. a, Yttrium film before hydrogenation. The material is metallic and, for the thickness used, virtually all the photons in the visible range are reflected. The mirror image of the knight is clearly visible. b, The yttrium film after approximately one minute in H_2 at a gas pressure of 0.9×10^5 Pa near the dihydride phase ($x \approx 1.8$). In agreement with measurements in bulk material^{18,19}, the film acts as a bandpass filter. In transmission it is dark brown, and in reflection it is blue as can be seen on the back side of the film where there is no reflection from the palladium protection layer. c, The film after further hydrogen uptake is in the trihydride γ -phase and has become highly transparent. (The faint knight mirror image is due to a residual reflection caused by the palladium overlayer). The transparency of the trihydride phase (c) is shown by the visibility of the chess-board pattern. The yellow colour indicates that the material remains transparent up to high photon energies. From absorption-edge measurements, we find that the gap energy $E_g \approx 1.8\text{ eV}$ with an absorption edge up

to 2.8 eV. At room temperature and 10^5 Pa hydrogen gas pressure, the transition from b to transparency occurs within a few seconds. Photograph c was taken after equilibrium was reached.

measurements with bulk resistivity data²⁰ suggests that this happens well within the β - γ coexisting phase region. The transmitted light intensity increases with time, as more h.c.p. YH_{3-6} phase precipitates. The abruptness with which the transition takes place, as well as the fact that we found a strong hysteresis between hydrogen ab- and de-sorption, both reflect the involvement of the first-order $\beta \rightarrow \gamma$ (f.c.c. $\rightarrow$ h.c.p.) phase transition.

A structural transition is not required for the occurrence of a metal-semiconductor transition. The electrical and optical transmission data given in Fig. 3 shows that the metal hydride LaH_x maintains its cubic symmetry over the whole concentration range ($2 < x < 3$) at temperatures above ~ 300 K (ref. 5). The data in Fig. 3 are similar to those in Fig. 2 although no transparency window (corresponding to the little peak around $t = 67$ s in Fig. 2) is observed near LaH_2 . Furthermore, the optical switching of LaH_x is clearly more gradual than for YH_x .

The electronic structure of the host metal is profoundly affected by the presence of hydrogen. This is especially true for the group III transition metals (Sc, Y, La) and the trivalent RE which can form dihydrides and trihydrides⁷. Despite a large and prolonged research effort, the nature of the ground state of metal trihydrides and the mechanism of the metal-insulator transition when the hydrogen concentration is increased from MH_2 to MH_3 has eluded clarification^{7,20,21}.

Photoemission data show that the dihydrides of Y, La, Ce and Pr are metallic whereas their trihydrides are non-metallic, and confirm that low-lying hybridized hydrogen bands are formed in

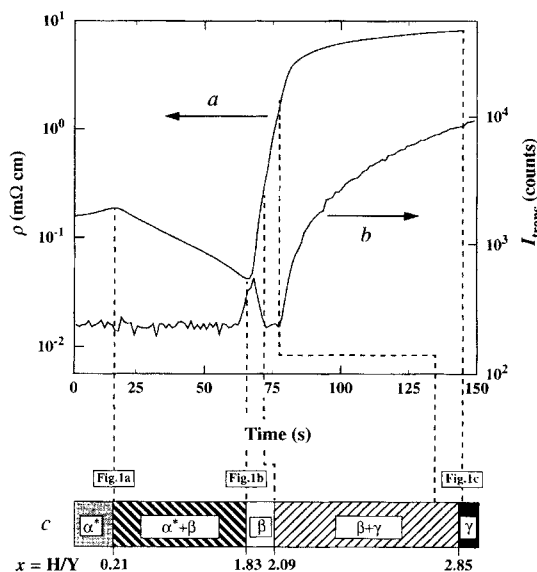


FIG. 2 *a*, Resistivity change during hydrogen loading ($p_{\text{H}_2} = 0.9 \times 10^5$ Pa) for a 500-nm yttrium film with a 20-nm Pd protective layer at room temperature as a function of time. The resistivity in the trihydride γ -phase is corrected for the effect of the palladium overlayer using a parallel resistor model. *b*, Influence of hydrogen on the light transmission intensity I_{trans} at 1.8 eV. The photon energy is chosen so as to have $E_g \approx \hbar\omega$, which implies a large influence of the band rearrangement due to octahedral occupancy in the superstoichiometric dihydride YH_{2+y} and a maximum transmission in the substoichiometric trihydride YH_{3-6} . After correction for the optical behaviour of the palladium layer over the total energy interval, we conclude that the yttrium trihydride film transmits nearly 100% of the incident light below this energy. This indicates that there is no free charge carrier reflection present below 1.8 eV and that the material is not metallic. *c*, Relation between the thermodynamic phases and the hydrogen concentration at room temperature as deduced from the experimentally determined phase diagram for the bulk yttrium-hydrogen system¹² and measurements on superstoichiometric yttrium dihydride²⁰. The dashed lines relate the resistivity and optical transmission to the various thermodynamic phases of the yttrium-hydrogen system. They also indicate when the photographs shown in Fig. 1*a-c* were taken.

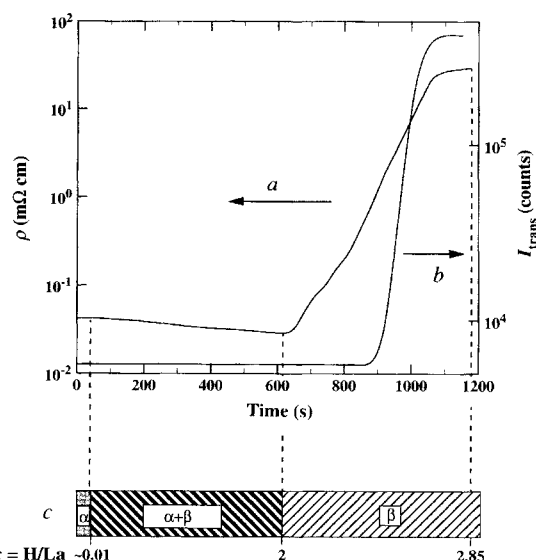


FIG. 3 As for Fig. 2*a-c* but for a 500-nm-thick lanthanum film with a 20-nm palladium protective layer. In contrast to YH_x , LaH_x remains cubic at temperatures above room temperature. There is thus no β - γ -phase miscibility gap above this temperature. The fact that a metal-semiconductor transition is also observed in LaH_x demonstrates that this transition is of electronic origin and not triggered by a structural phase transformation.

the trihydrides 3–6 eV below the metal d -states⁷. They do not allow us to determine, however, whether the ground state of the trihydride is semiconducting or semimetallic. So far, the magnitude of the gap in the electronic structure of the trihydrides is unknown^{18,19}. In the literature it has been noted that trihydrides are dark-blue to black⁶, grey or deep purple black²² or light purple²³.

Band-structure calculations, although very important for the understanding of the basic features of the metalhydrides, have been unable to predict the magnitude of the semiconducting gap. In YH_3 , which is the simplest trihydride in the sense that yttrium has no occupied f -states and is relatively light (which minimizes spin-orbit effects), Switendick²⁴ found a gap, $E_g \approx 1.5$ eV, whereas recently in self-consistent band-structure calculations Dekker *et al.*²⁵ and Wang and Chou²⁶ found that YH_3 is a metal when in the hexagonal HoD_3 structure, and at best a semimetal when in a modified hexagonal structure in which the hydrogen atoms on formerly octahedral sites have been moved towards the yttrium hexagonal planes. To reconcile this with the strong increase in resistivity when x is increased from 2 to 3, Wang and Chou²⁶ proposed that the perfect nesting of the hole and electron pockets at the centre of the Brillouin zone could lead to an excitonic insulator and to Peierls distortions.

However, the very small optical absorption coefficient α (below the absorption edge) found here clearly demonstrates that there is no free charge carrier reflection, thus confirming the existence of a semiconducting state. We have found the following additional facts in favour of a semiconducting ground state. (1) The low charge carrier density (10^{19} cm^{-3}) determined from Hall-effect measurements in samples with typically 2×10^{21} impurities per cm^{-3} (corresponding to $\text{YH}_{2.9}$). In the case of a two-band solid, one would expect a much higher carrier density. (2) The negative temperature coefficient of the electrical resistivity ($d\rho/dT$), for $2 \text{ K} < T < 350 \text{ K}$. (3) The further increase of the resistivity when x is increased to near-stoichiometry (up to $x = 3$) under high pressures (several GPa H_2 pressure). (4) The weak magnetoresistance in fields up to 7 T at temperatures between 4.2 and 300 K.

The failure of the one-electron band-structure calculations to predict a semiconducting gap in YH_3 (in fact all existing self-consistent calculations predict a substantial band overlap) is a

strong indication of the importance of electron–electron correlation effects²⁷.

We have shown that one can continuously and reversibly move through the metal–insulator transition by simply changing the hydrogen to yttrium ratio from 1.8 to 2.9 in yttrium thin films, covered by a palladium overlayer. This can be done in a remarkably short time by just changing the pressure of the hydrogen gas surrounding the sample at room temperature. The spectacular changes in optical properties (from a shiny mirror to a yellow transparent window) could potentially lead to technological applications of metal–hydride films because: (1) P. H. L. Notten *et al.* (manuscript in preparation) have found that our samples can easily be charged electrolytically and (2) M. Ouwkerk *et al.*,

(manuscript in preparation) found in the hydrides of some RE alloys optical switching times much faster than those of YH₃ and LaH₃. We have determined the direct energy gap of YH₃ (1.8 eV); its magnitude disagrees with recent band-structure calculations^{25,26} that predict a (semi) metallic ground state, and is far too large for an excitonic insulator²⁸. Our results strongly suggest that electron correlation effects are important in YH_x, LaH_x and REH_x.

Because the electrical and optical phenomena are not limited to yttrium and lanthanum, but are also found in rare earths and their alloys, the technique used in this work offers many possibilities of unravelling the intriguing properties of trihydrides and non-stoichiometric higher hydrides, MH_x, in general. □

Received 6 October 1995; accepted 1 February 1996.

- Mott, N. F. in *Metal–Insulator transitions* 2nd edn (Taylor & Francis, London 1990).
- Verleur, H. W., Barker, A. S. Jr & Berglund, C. N. *Phys. Rev.* **172**, 788–798 (1968).
- Libowitz, G. G. *Ber. Bunsenges. phys. Chem.* **77**, 837–845 (1973).
- Libowitz, G. G. & Maeland, A. J. in *Handbook on the Physics and Chemistry of Rare Earths* (eds Gschneidner, K. A. & Eyring, L.) 299–336 (North Holland, Amsterdam, 1979).
- Vajda, P. in *Handbook on the Physics and Chemistry of Rare Earths* Vol. 20 (eds Gschneidner, K. A. & Eyring, L.) 207–291 (Elsevier, Amsterdam, 1995).
- Mueller, W. M., Blackledge, J. P. & Libowitz, G. G. in *Metal Hydrides* (eds Mueller, W. M., Blackledge, J. P. & Libowitz, G. G.) Ch. 9 & 10 (Academic, New York, 1968).
- Schlapbach, L. & Gupta, M. in *Hydrogen in Intermetallic Compounds* (ed. Schlapbach, L.) 139–217 (Topics in Appl. Phys. Vol. 63, Springer, Berlin, 1988).
- Bonnet, J. E., Juckum, C. & Lucasson, A. *J. Phys. F: Met. Phys.* **12**, 699–711 (1982).
- Pebler, A. & Wallace, W. A. *J. phys. Chem.* **66**, 148–151 (1962).
- Lundin, C. E. & Blackledge, J. P. *J. Electrochem. Soc.* **109**, 838–842 (1962).
- Plotow, H. E., Osborne, D. W. & Otto, K. J. *chem. Phys.* **36**, 866–872 (1962).
- Yannopoulos, L. N., Edwards, R. K. & Wahlbeck, P. G. *J. phys. Chem.* **69**, 2510–2515 (1965).
- Alfeld, G. *Ber. Bunsenges. phys. Chem.* **76**, 746–755 (1972).
- Buccur, R. V. & Flanagan, T. B. *J. phys. Chem.* **88**, 225–241 (1974).
- Frazier, G. A. & Glosser, R. J. *J. Less-Common Met.* **74**, 89–96 (1980).

- Lanford, W. A. *Nuc. Instrum. Meth.* **149**, 1–8 (1978).
- Granqvist, C. G. in *Handbook of Inorganic Electrochromic Materials* (ed. Granqvist, C. G.) Ch. 8, 9, 11 (Elsevier, Amsterdam, 1995).
- Weaver, J. H., Rosel, R. & Peterson, D. T. *Phys. Rev.* **B19**, 4855–4866 (1979).
- Peterman, D. J., Weaver, J. H. & Peterson, D. T. *Phys. Rev.* **B23**, 3903–3913 (1981).
- Vajda, P. & Daou, J. N. *Phys. Rev. Lett.* **66**, 3176–3178 (1991).
- Shinar, J., Dehner, B., Barnes, R. G. & Beaudry, B. J. *Phys. Rev. Lett.* **64**, 563–566 (1990).
- Osterwalder, J. Z. *Phys. Rev. Lett.* **61**, 113–128 (1985).
- Braccioni, P., Pörschke, E. & Lässer, R. *Appl. Surf. Sci.* **32**, 392–408 (1988).
- Switendick, A. C. *Int. J. Quant. Chem.* **5**, 459–470 (1971).
- Dekker, J. P., van Ek, J., Lodder, A. & Hulberts, J. N. J. *Phys.: Cond. Mat.* **5**, 4805–4816 (1993).
- Wang, Y. & Chou, M. Y. *Phys. Rev. Lett.* **71**, 1226–1229 (1993).
- Zaanen, J., Sawatzky, G. A. & Allen, J. W. *Phys. Rev. Lett.* **55**, 418–421 (1985).
- Halperin, B. I. & Rice, T. M. in *Solid State Physics* Vol. 21 (eds Seitz, F., Turnbull, D. & Ehrenreich, H.) 115–192 (Academic, New York, 1968).

ACKNOWLEDGEMENTS. We acknowledge discussions with G. Sawatzky, M. Kremers, P. H. L. Notten, P. A. Duine, H. van Houten, J. van de Ven, M. Ouwkerk, F. J. A. den Broeder and P. J. Kelly. We thank B. Hjörvarsson, S. Olafsson and E. Karlsson for their help with the ¹⁵N hydrogen profiling measurements. This work is part of the research programme of the Stichting voor Fundamenteel Onderzoek der Materie (FOM) which is supported by NWO.

Oxygen-isotope record of sea level and climate variations in the Sulu Sea over the past 150,000 years

Braddock K. Linsley*†

* Department of Geology and Geophysics, Rice University, Houston, Texas 77251, USA

THE Sulu Sea is located in the 'warm pool' of the western Pacific Ocean, where mean annual temperatures are the highest of anywhere on Earth. Because this large heat source supplies the atmosphere with a significant portion of its water vapour and latent heat, understanding the climate history of the region is important for reconstructing global palaeoclimate and for predicting future climate change. Changes in the oxygen isotope composition of planktonic foraminifera from Sulu Sea sediments have previously been shown to reflect changes in the planetary ice volume at glacial–interglacial and millennial timescales, and such records have been obtained for the late Pleistocene epoch and the last deglaciation^{1–3}. Here I present results that extend the millennial time resolution record back to 150,000 years before present. On timescales of around 10,000 years, the Sulu Sea oxygen-isotope record matches changes in sea level deduced from coral terraces on the Huon peninsula⁴. This is particularly the case during isotope stage 3 (an interglacial period 23,000 to 58,000 years ago) where the Sulu Sea oxygen-isotope record deviates from the SPECMAP deep-ocean oxygen-isotope record⁵. Thus these results support the idea^{4,6} that there were higher sea levels

and less continental ice during stage 3 than the SPECMAP record implies and that sea level during this interglacial was just 40–50 metres below present levels. The subsequent rate of increase in continental ice volume during the return to full glacial conditions was correspondingly faster than previously thought.

The oxygen isotope ($\delta^{18}\text{O}$) data presented here are from Ocean Drilling Program (ODP) sites 769 and 768 in the Sulu Sea (Table 1, and Figs 1, 2 and 3). The $\delta^{18}\text{O}$ of the planktonic foraminifera *Globigerinoides ruber* was measured at ~400-yr resolution to develop a detailed palaeoclimate history of this region. Past work has shown that high-resolution sediment archives have been preserved at certain locations in the Sulu Sea owing to relatively high sediment accumulation rates^{1–3} and reduced benthic mixing in the dysaerobic conditions of the deep basin⁷.

The high quality and resolution of the Site 769 record is reflected in the preservation of both the full glacial–interglacial ice-volume amplitude and several millennial-scale climate events also found in higher-latitude palaeoclimate records. The glacial–interglacial $\delta^{18}\text{O}$ amplitude in this record (and other Sulu Sea records^{3,8}) is only 1.3‰, almost equalling the 1.25‰ planetary ice-volume signal⁹. This indicates that the Sulu Sea experienced no significant change in sea surface temperatures (SSTs) between glacial and interglacial times, a conclusion also in agreement with recent foraminifera transfer function results from the western tropical Pacific which show that SST differed by less than 2 °C from the present¹⁰.

The potentially most significant aspect of the Site 769 isotope record is its implications for eustatic sea level and planetary ice volume during interglacial isotope stage 3 (~23 to ~58 kyr ago). Throughout stage 3, relative $\delta^{18}\text{O}$ values at Site 769 are 0.5‰ more negative than the SPECMAP⁵ composite deep-ocean $\delta^{18}\text{O}$ record which is thought to reflect global ice-volume changes (Fig. 3a). Instead, during stage 3, and throughout the 150-kyr record, Sulu Sea planktonic $\delta^{18}\text{O}$ is closely correlated with the sea-level record derived from Huon peninsula coral terraces⁴ (Fig. 3a). This correlation, in conjunction with the 1.3‰ glacial–interglacial $\delta^{18}\text{O}$ amplitude, suggests that the lower-frequency components

† Present address: Department of Geological Sciences, University at Albany, State University of New York, Albany, New York 12222, USA.

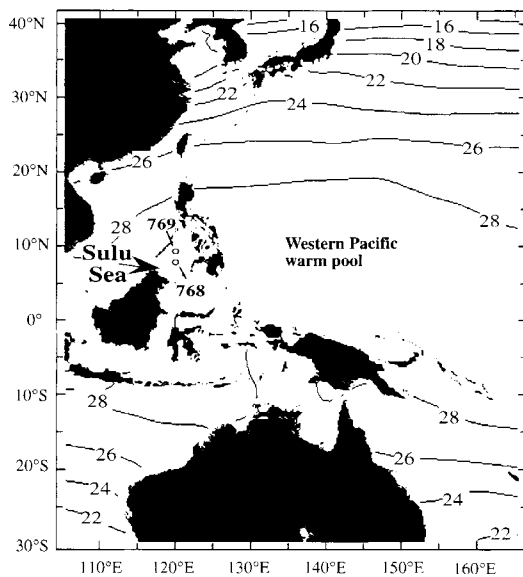


FIG. 1 Map showing the location of Ocean Drilling Program (ODP) sites 769 and 768 in the Sulu Sea. Annually averaged sea surface temperature (SST) contours (in °C)³² are also shown, highlighting the area of the western Pacific warm pool. In the Sulu Sea, modern SSTs average ~28.5 °C and have an annual range from 26.5 to 30.5 °C. The basin is surrounded by a shallow sill which is mostly <100 m deep. Today, there is an open exchange of surface water between the Sulu Sea and surrounding seas^{32,33} over the shelf and through channels (200–400 m deep) that connect the basin to the South China and Celebes seas. SST fields in the Sulu Sea indicate that the coldest waters occur in winter months (November–February), the result of through-flow across the Philippine archipelago from the open western Pacific^{32,33}. Although Pleistocene sea level low stands would have exposed much of the sill surrounding the basin, previous studies have shown that the basin was never completely isolated^{1,2}. Site 769 was cored above the basin's carbonate lysocline⁸ at 3,643 m on an isolated bathymetric high, and Site 768 was cored at 4,300 m in the Sulu Sea central deep basin.

of this $\delta^{18}\text{O}$ record are primarily a function of global ice-volume and sea-level variations.

The agreement between Site 769 $\delta^{18}\text{O}$ and the Huon terrace estimates of sea level during stage 3 (Fig. 3a) is particularly interesting, as these two records indicate significantly less global ice volume than does the deep-ocean $\delta^{18}\text{O}$ signal, even when this last signal is corrected for suspected cooler temperatures⁶. The relatively low $\delta^{18}\text{O}$ recorded by planktonic foraminifera in the Sulu Sea during stage 3 is apparently unique, and has not been reported in other western tropical Pacific locations such as the Banda Sea¹¹ or South China Sea¹². However, both of these basins are more open to through-flow from the western Pacific. Perhaps the semi-isolated configuration of the basin has insulated the Sulu Sea from oceanographic changes, preserving the true ice-volume/sea-level signal. Alternatively, Sulu Sea surface waters during stage 3 could have become warmer and/or have had lower salinity with respect

to the deep ocean and ice-volume signal. But the close correspondence between the sea-level curve derived from the Huon peninsula coral terraces⁴ and Site 769 $\delta^{18}\text{O}$, coupled with the 1.3‰ glacial–interglacial $\delta^{18}\text{O}$ amplitude, argues against this and suggests that Site 769 $\delta^{18}\text{O}$ is a good proxy for sea-level changes with little effect of changes in local SST and/or $\delta^{18}\text{O}$ of sea water. Although the dating of some Huon corals may be affected by diagenetic alteration¹³, the relative terrace elevations are accurate. In agreement with these results, further evidence for relatively high sea levels during stage 3 comes from submerged speleothems in the Bahamas¹⁴.

In addition to the ice-volume/sea-level component, the Sulu Sea $\delta^{18}\text{O}$ values document several pronounced millennial-scale oceanographic events that reflect climate changes in the warm pool. Both cores 769 and 768 show 0.4‰ $\delta^{18}\text{O}$ decreases just before and after the Younger Dryas chronozone (YD), indicating the basin-wide coherence of the planktonic $\delta^{18}\text{O}$ signal (Table 1 and Fig. 3b). The radiocarbon-dated portions of both sites 769 and 768 show that the largest deglacial $\delta^{18}\text{O}$ shift correlates with the deep ocean and Northern Hemisphere records (Table 1 and Fig. 3), in contrast to the results from the Byrd Antarctic ice core which indicate that deglacial warming in Antarctica began ~3,000 yr before the onset of the warming in Greenland¹⁵. This suggests that the Sulu Sea was responding primarily to Northern Hemisphere climate changes at this time. Site 769 $\delta^{18}\text{O}$ also reveals other abrupt climate shifts that closely match those in Greenland and Antarctic ice cores (Fig. 3b)^{16,17}. Most notably, similar events are observed at the isotope stage 5/4 glacial inception, and during interglacial stage 5e (Eemian). In addition, the $\delta^{18}\text{O}$ values from Site 769 show a pronounced 'step' at the stage 6/5 glacial transition, similar to that observed during the YD in this core.

A question posed by the presence of abrupt deglacial climate events in the Sulu Sea that are close in age to previously identified events in the North Atlantic^{17–20} and high latitudes^{20–22} is whether these abrupt $\delta^{18}\text{O}$ events reflect local changes in surface ocean conditions or, like the long-term trend, reflect global changes in the $\delta^{18}\text{O}$ of sea water. The overall $\delta^{18}\text{O}$ record presented here provides some clues. If the large-scale glacial–interglacial climate changes of the past 150 kyr did not produce significant SST changes in the Sulu Sea, it seems unlikely that lower-amplitude climate changes, such as the YD, could reflect lowered SST. Instead, it seems more likely that this Sulu Sea $\delta^{18}\text{O}$ record is a passive recorder of these climate signals. One potential mechanism that has been suggested previously, and which would explain these deglacial events, is advection along the sea surface (or through the atmosphere) of ^{18}O -depleted melt water from higher latitudes^{23,24}.

To explain the abrupt increases in planktonic $\delta^{18}\text{O}$ at the 5/4 boundary and during interglacial stage 5e, (a) different mechanism(s) must be invoked because ice sheets were greatly reduced in size or growing at these times. One possibility is that these two events may be related to oceanographic changes due to volcanism. I note that the large-scale Toba eruption in Sumatra occurred at ~75 kyr, at the stage 5/4 climate boundary²⁵. This is also the same

FIG. 2 Oxygen isotope analysis ($\delta^{18}\text{O}$ values) of the surface-dwelling planktonic foraminifera *Globigerinoides ruber* from ODP Site 769, plotted against depth in core. The upper 10 m of Site 769 was sampled at ~3-cm intervals for this study. Methods for the isotopic analyses are presented elsewhere^{1,2}. Numbers indicate marine isotope stages. Vertical arrows mark the position of accelerator mass spectrometry (AMS) radiocarbon ages shown in Table 1. (Here $\delta^{18}\text{O} = [(^{18}\text{O}/^{16}\text{O})_{\text{sample}} / (^{18}\text{O}/^{16}\text{O})_{\text{standard}}] - 1$, where the PDB standard is used.)

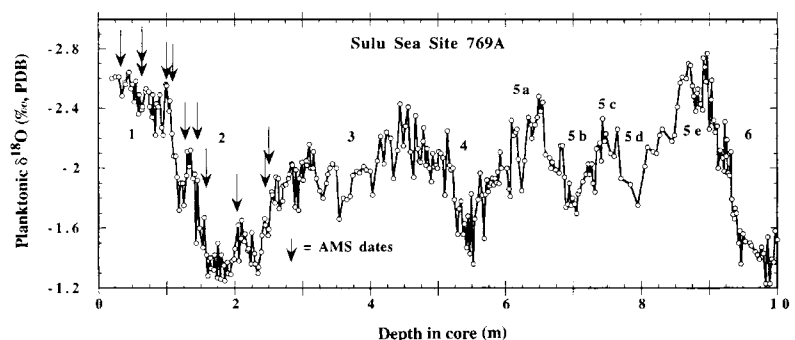
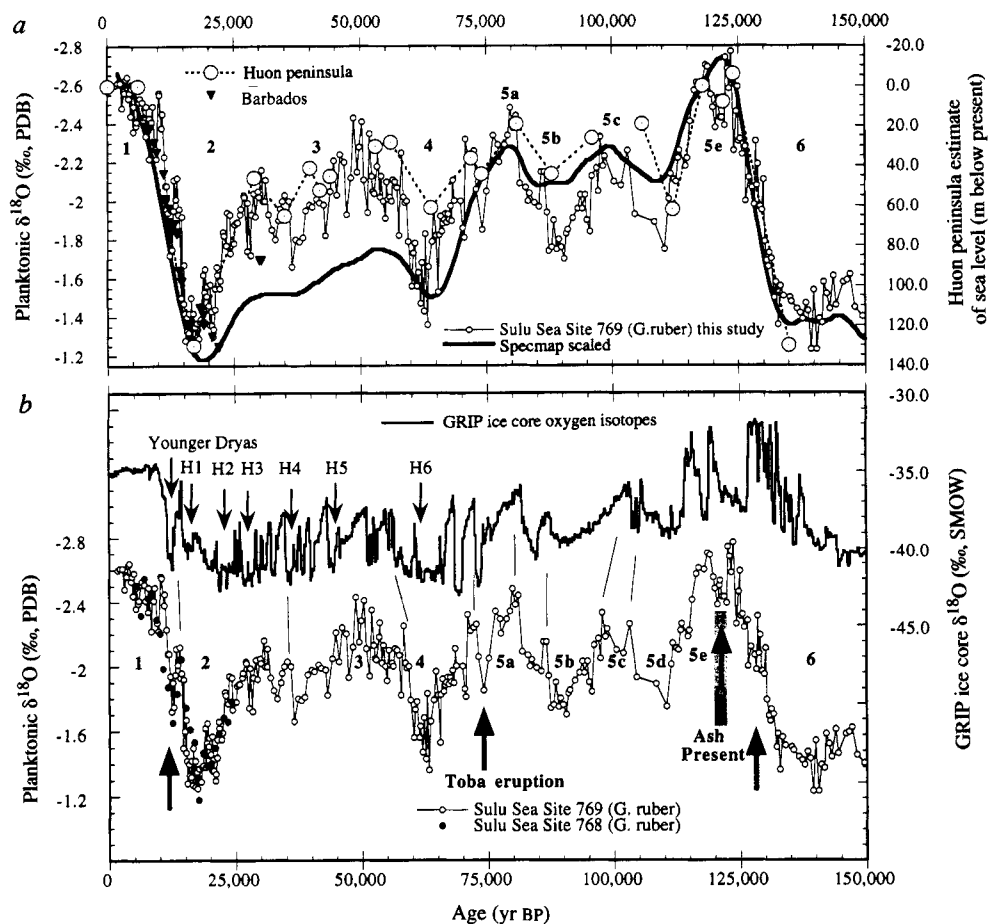


FIG. 3 a, ODP Site 769 planktonic $\delta^{18}\text{O}$ results (thin solid line with open circles) compared to a scaled version of the SPECMAP deep-ocean $\delta^{18}\text{O}$ record (bold solid line), and past sea level estimated from the Huon peninsula⁴ (dashed line, large open circles) and Barbados^{9,34} (filled triangles). Numbers indicate isotope stages. Note the close correspondence between the Huon peninsula sea-level record and Site 769 $\delta^{18}\text{O}$, in particular during stage 3 when the SPECMAP record is $\sim 0.5\%$ less negative. See text for discussion. b, Lower curve, oxygen isotope ($\delta^{18}\text{O}$) analyses from Sites 769 (thin line, open circles) and 768 (filled circles) plotted against time based on the age model given in Table 1. Large arrows mark locations of intervals discussed in the text, and numbers indicate the marine isotope stages. Upper curve, GRIP Greenland summit ice-core $\delta^{18}\text{O}$ results spanning the past 150,000 yr (ref. 17). The temporal position of the Younger Dryas and Heinrich events H1–H6 are marked by small arrows.



time as North Atlantic deep-water reorganization²⁶, a decrease in arboreal pollen in the Grande Pile peat record²⁷, and an increase in Vostok ice core δD (ref. 28). Evidence for stage 5e volcanism near the Sulu Sea comes from a minor amount of fine-grained sedimentary volcanic ash (63–250 μm) that is present during the

brief interval of higher $\delta^{18}\text{O}$ values at ~ 122 kyr in Site 769 (Fig. 3b). This event suggests climate instability during the last interglacial in the western Pacific, as indicated by some (but not all) higher-latitude palaeoclimate records^{17,27,29}.

The lack of well defined events of lower $\delta^{18}\text{O}$ in the Sulu Sea

during stage 3 may indicate that the volume of ^{18}O -depleted melt water associated with the stage 3 Dansgaard–Oeschger oscillations¹⁹ was less than the volume of melt water released during deglacial events at the stage 2/1 and 6/5 boundaries. This interpretation also agrees with the evidence (from the Sulu Sea $\delta^{18}\text{O}$ values and the Huon peninsula sea-level record) of higher sea levels and less continental ice during stage 3 than implied by the SPECMAP deep-ocean $\delta^{18}\text{O}$ record. In turn, the Site 769 results independently corroborate the sea-level record derived from the Huon coral terraces⁴ and support deep-ocean cooling and/or salinity effects on open-ocean benthic $\delta^{18}\text{O}$ records during stage 3 (ref. 6). These results also imply that continental ice build-up to full glacial conditions following stage 3 occurred at a higher rate, and within only a period

TABLE 1 Age models for sites 769 and 768

ODP site	Depth in core (cm)	AMC ^{14}C age (^{14}C yr BP)	Error*	Corrected age (cal. yr. BP)†	Species	SPECMAP age (cal. yr. BP)
769	0.31	3,180	60	2,613	<i>G. sacculifer</i>	—
769	0.62	5,430	65	5,178	<i>G. ruber</i>	—
769	0.62 dup.	5,515	70	5,275	<i>G. sacculifer</i>	—
769	1.01	9,550	95	10,422	<i>G. sacculifer</i>	—
769	1.07	10,120	80	11,151	<i>G. ruber</i>	—
769	1.28	11,500	185	12,901	<i>G. ruber</i>	—
769	1.45	12,950	135	14,716	<i>G. sacculifer</i>	—
769	1.6	13,450	110	15,336	<i>G. sacculifer</i>	—
769	2.05	16,400	135	18,935	<i>G. ruber</i>	—
769	2.45	18,720	155	21,694	<i>G. ruber</i>	—
769	2.5	19,275	150	22,345	<i>G. sacculifer</i>	—
769	5.2	—	—	—	—	59,000
769	6.1	—	—	—	—	71,000
769	6.5	—	—	—	—	80,000
769	7.5	—	—	—	—	99,000
769	7.9	—	—	—	—	110,000
769	8.5	—	—	—	—	115,000
769	9.1	—	—	—	—	126,000
768	0.85	9,740	80	10,665	<i>G. sacculifer</i>	—
768	1.05	12,525	80	14,187	<i>G. sacculifer</i>	—

* Error given is 1σ .

† Calendar year calculated following ref. 30 for raw uncorrected ages $< 9,000$ ^{14}C yr, and following ref. 31 for raw ages $> 9,000$ ^{14}C yr. Sedimentation rates for the radiocarbon-dated portion of the Site 769 record average 11.2 cm kyr⁻¹, whereas over the entire record sedimentation rates average 7.5 cm kyr⁻¹.

of ~10 kyr, much more rapidly than the deep-ocean $\delta^{18}\text{O}$ record indicates. □

Received 31 May 1995; accepted 14 February 1996.

1. Linsley, B. K. & Thunell, R. C. *Paleoceanography* **5**, 1025–1039 (1990).
2. Linsley, B. K. & Dunbar, R. B. *Paleoceanography* **9**, 317–340 (1994).
3. Kudrass, H. R., Eriksen, H., Vollbrecht, R. & Weiss, W. *Nature* **349**, 406–409 (1991).
4. Chappell, J. & Shackleton, N. J. *Nature* **324**, 137–140 (1986).
5. Imbrie, J. et al. in *Milankovitch and Climate* (eds Berger, A. et al.) 269–305 (Reidel, Norwell, MA, 1984).
6. Shackleton, N. J. *Quat. Sci. Rev.* **6**, 183–190 (1987).
7. Kuehl, S. A., Fuglseth, T. J. & Thunell, R. C. *Mar. Geol.* **111**, 15–35 (1993).
8. Linsley, B. K., Thunell, R. C., Morgan, C. & Williams, D. F. *Mar. Micropaleon.* **9**, 395–418 (1985).
9. Fairbanks, R. G. *Nature* **342**, 637–642 (1989).
10. Thunell, R. C., Anderson, D., Gellar, D. & Miao, Q. *Quat. Res.* **41**, 255–264 (1994).
11. Ahmad, S. M., Guichard, F., Hardjavidjaksana, K., Adisaputra, M. K. & Labeyrie, L. D. *Mar. Geol.* **22**, 385–397 (1995).
12. Luejia, W. & Pinxian, W. *Paleoceanography* **5**, 77–90 (1990).
13. Gallap, C., Edwards, R. L. & Johnson, R. G. *Science* **263**, 796–800 (1994).
14. Richards, D. A., Smart, P. L. & Edwards, R. L. *Nature* **367**, 357–360 (1994).
15. Sowers, T. & Bender, M. *Science* **269**, 210–214 (1995).
16. Bender, M. et al. *Nature* **332**, 663–666 (1994).
17. Greenland Ice-Core Project (GRIP) Members *Nature* **364**, 203–207 (1993).
18. Bond, G. et al. *Nature* **365**, 143–147 (1993).
19. Dansgaard, W. et al. *Nature* **364**, 218–220 (1993).
20. Broecker, W. S. *Nature* **372**, 421–424 (1994).
21. Mathewes, R. W., Heusser, I. E. & Patterson, R. T. *Geology* **21**, 101–104 (1993).
22. Denton, G. H. & Hendy, C. H. *Science* **264**, 1434–1437 (1994).
23. Duplessy, J. C. et al. *Earth planet. Sci. Lett.* **103**, 27–40 (1991).
24. Anderson, D. M. & Thunell, R. C. *Quat. Sci. Rev.* **12**, 465–473 (1993).
25. Ninkovitch, D., Shackleton, N. J., Abek-Monem, A. A., Obradovitch, J. D. & Izett, G. *Nature* **276**, 574–577 (1978).
26. Keigwin, L. D. & Jones, G. A. *J. geophys. Res.* **99**, 12397–12410 (1994).
27. Thouveny, N. et al. *Nature* **371**, 503–506 (1994).
28. Jouzel, J. et al. *Nature* **329**, 403–408 (1987).
29. Field, M. H., Huntley, B. & Muller, H. *Nature* **371**, 779–783 (1994).
30. Stuiver, M. & Braziunas, T. F. *Radiocarbon* **35**, 137–189 (1993).
31. Bard, E., Arnold, M., Fairbanks, R. G. & Hamelin, B. *Radiocarbon* **35**, 191–199 (1993).
32. Levitus, S. & Boyer, T. P. *NOAA Atlas NESDIS 4* (US Dept of Commerce, US Govt Printing Office, Washington DC, 1994).
33. Wyrtki, K. *NAGA-Rep. No. 2* (Univ. California, La Jolla, 1961).
34. Bard, E., Hamelin, B. & Fairbanks, R. G. *Nature* **346**, 456–458 (1990).

ACKNOWLEDGEMENTS. I thank R. B. Dunbar for guidance and use of the stable-isotope facility at Rice University and R. C. Thunell for initiating earlier Sulu Sea research. This work was supported by the US National Science Foundation.

Constraints from partitioning experiments on the composition of subduction-zone fluids

Hans Keppler

Bayerisches Geoinstitut, Universität Bayreuth, 95440 Bayreuth, Germany

THE generation of calc-alkaline magmas in subduction zones is thought to be the most important mechanism for the growth of continental crust since the Proterozoic eon. It is widely assumed that most of these magmas are products of fluid-triggered melting in the mantle wedge above the subducted slab^{1,2}. Fluid transport from the subducted slab into the zone of melting has also been invoked in order to explain many of the trace element and radiogenic isotope characteristics of calc-alkaline magmas^{3–7}. Here I report experimental data on the partitioning of trace elements between fluids, silicate melts and minerals, which suggest that the agent responsible for the transport of trace elements in subduction zones may be an alkali-chloride-rich aqueous fluid. The data show that chemical transport by such a fluid can generate the trace element and isotope enrichment pattern typical for calc-alkaline magmas, including the enrichment of large ionic lithophile elements, lead and uranium, and the characteristic depletion in niobium and tantalum.

Figure 1a shows the type of trace-element enrichment pattern characteristic of subduction-zone volcanics. While the elements below the dashed line are slightly depleted relative to the abun-

dances in mid-ocean-ridge basalts (MORBs), others are strongly enriched. In Fig. 1a, the elements are ordered in such a way that the compatibility during normal, fluid-absent melting increases to the right. The fact that the relative enrichment of elements does not correlate with this sequence means that it cannot be explained by a simple melting event involving only fractionation of elements between a silicate melt and solid minerals. Rather, fluid transport may be responsible for the observed pattern.

The most important constituent of the subduction zone fluids is certainly water^{1,2}. However, oceanic sediments and altered basalts in the subducted slab have been in contact with sea water and therefore contain abundant chloride. As chloride readily substitutes for OH groups in minerals such as amphibole, it can be transported to the depth of amphibole dehydration^{1,2} where it will be released as a component of the hydrous fluid. Two lines of evidence support the presence of a chloride-rich hydrous fluid in subduction zones: the strongly elevated chlorine concentrations in calc-alkaline magmas (460–2,000 p.p.m.; ref. 6) and the presence of highly saline fluid inclusions in subduction zone metasediments from Dora Maira⁸. Most importantly, numerous experimental studies have shown that the presence of chloride strongly increases fluid/mineral and fluid/melt partition coefficients for many trace elements, owing to the formation of ion pairs and undissociated complexes in supercritical solutions^{9,10}. Using the concept of 'hard' and 'soft' bases and acids¹¹, one can even predict which elements should be preferentially mobilized by a chloridic fluid. Chloride, as a moderately hard base, should preferentially react with moderately hard acids, such as Rb^+ , Ba^{2+} , Sr^{2+} , Pb^{2+} . On the other hand, extremely hard acids, such as Nb^{5+} and Ta^{5+} , are not expected to react with chloride and will therefore not be transported by the fluid.

To test whether chemical transport by pure water or by a chloride-rich fluid can explain the trace element abundances in subduction-zone volcanics, experimental data on the partitioning of trace elements between such fluids and the minerals of the slab and the mantle wedge are necessary. Direct measurements of such partition coefficients are very difficult^{9,12} or impossible. The extremely small diffusion coefficients of many trace elements in solid minerals often make it impossible to attain equilibrium within realistic run durations. Another problem in direct measurements of fluid/melt partition coefficients is caused by the recrystallization and sometimes incongruent dissolution of the solids during the experiment. Recrystallization will almost always lead to the formation of fluid inclusions in the mineral studied. If a trace element is highly incompatible in a mineral, even traces of submicroscopic fluid inclusions can lead to errors of several orders of magnitude in the measured partition coefficients. Incongruent dissolution can lead to the formation of tiny amounts of accessory phases that strongly concentrate certain trace elements and complicate the interpretation of the experiments. Therefore, in the present study, the fluid/mineral partition coefficients ($D^{\text{fluid/mineral}}$) have been determined indirectly by measuring the partitioning of trace elements between a fluid and a silicate melt. Because this measurement involves two liquid phases, attainment of equilibrium is not a problem. The fluid/mineral partition coefficients can then be estimated using published mineral/melt partition coefficients^{13,14,20–22} according to the equation $D^{\text{fluid/mineral}} = D^{\text{fluid/melt}} / D^{\text{mineral/melt}}$. Strictly, this equation only holds if the melt composition studied is in equilibrium with the mineral of interest. The most important carrier of incompatible elements both in the subducted oceanic crust and in the mantle wedge is clinopyroxene. Therefore, the partitioning of trace elements between hydrous fluids and an andesitic melt was measured at 1,040–1,100 °C and 3–20 kbar, as under these conditions, this melt is very close to saturation with clinopyroxene¹⁵. Although the melt is not simultaneously saturated with olivine, orthopyroxene, garnet and amphibole, the data can also be used to at least approximately determine the $D^{\text{fluid/mineral}}$ for these phases.

All experiments were carried out in Pt or Pt–5% Rh capsules with an initial weight ratio of melt to fluid of 1:1. Either pure H_2O

TABLE 1 Trace element partitioning between andesitic melt and fluid, and calculated fluid/clinopyroxene partition coefficients

Element	Experiment no.	Duration (days)	Fluid (p.p.m.)	Melt (p.p.m.)	$D^{\text{fluid/melt}}$	$D^{\text{cpx/melt}}$	Reference	$D^{\text{fluid/cpx}}$
Chloride-free hydrous fluid								
K	F153	5	< 300	12,000 (0.1)	< 0.02	0.0072	20	< 3
Rb	F149	5	2,167 (3)	6,793 (0.4)	0.32 (3)	0.002	21	160 (3)
Sr	F151	5	12 (10)	740 (1)	0.016 (11)	0.13	20	0.12 (11)
Ba	F151	5	29 (10)	924 (1)	0.031 (11)	0.00068	20	46 (11)
La	F151	5	< 10	587 (1)	< 0.02	0.054	20	< 0.4
Nb	F149	7	< 3	713 (0.1)	< 0.004	0.0077	20	< 0.5
						0.06	13	< 0.07
Ti	F153	5	24 (1)	911 (0.3)	0.026 (1)	0.384	20	0.068 (1)
U	F150	7	< 10	2,962 (1)	< 0.003	0.01	22	< 0.3
Pb	F150	7	287 (1)	3,263 (1)	0.088 (2)	0.072	20	1.2 (2)
Th	F153	5	49 (2)	2,103 (1)	0.023 (3)	0.003	22	7.7 (3)
Zn	F153	5	12 (1)	635 (1)	0.019 (2)	1	Estimated	0.019 (2)
5 m (Na,K)Cl fluid								
K	F132	20	80,000 (0.5)	30,000 (0.5)	2.7 (1)	0.0072	20	370 (1)
Rb	F131	20	6,521 (6)	2,460 (6)	2.7 (12)	0.002	21	1,300 (12)
Sr	F133	20	186 (0.6)	679 (0.6)	0.27 (1)	0.13	20	2.1 (1)
Sr*	F146	7	296 (0.2)	528 (0.2)	0.56 (0.4)	0.13	20	4.3 (0.4)
Ba	F133	20	245 (0.2)	790 (0.3)	0.31 (0.5)	0.00068	20	460 (0.5)
Ba*	F146	7	305 (0.4)	549 (0.8)	0.55 (1)	0.00068	20	809 (1)
La	F133	20	45 (1)	802 (0.7)	0.056 (2)	0.054	20	1.0 (2)
La*	F146	7	42 (1.4)	792 (0.1)	0.053 (1)	0.054	20	1.0 (1)
Gd	F133	20	49 (2)	943 (0.8)	0.052 (3)	0.36	20	0.14 (3)
Lu	F133	20	51 (0.4)	1,094 (0.1)	0.047 (0.5)	0.43	20	0.11 (0.5)
Nb	F131	20	< 3	764 (0.5)	< 0.005	0.0077	20	< 0.6
						0.06	13	< 0.08
Nb*	F147	7	< 5	2,554 (0.9)	< 0.002	0.0077	20	< 0.3
Ta	F131	20	< 4	1,013 (1.5)	< 0.004	Similar Nb		Similar Nb
Ti	F137	11	< 5	878 (0.2)	< 0.005	0.384	20	< 0.01
Zr	F137	11	< 6	1,230 (0.7)	< 0.005	0.123	20	< 0.04
U	F137	11	1,321 (2)	3,671 (0.5)	0.36 (2)	0.01	22	36 (2)
U*	F147	7	1,628 (0.8)	2,343 (0.5)	0.69 (1)	0.01	22	69 (1)
Pb	F137	11	3,204 (0.7)	760 (1.2)	4.2 (2)	0.072	20	58 (2)
Th	F137	11	30 (3)	2,409 (0.1)	0.012 (3)	0.003	22	4 (3)
Zn	F137	11	587 (0.7)	77 (0.8)	7.6 (1)	1	Estimated	7.6 (1)
B	F142	7	101 (0.4)	92 (0.3)	1.1 (0.7)	0.12	14	9.1 (0.7)
Be	F137	11	0.04 (10)	291 (0.5)	0.0001 (10)	—	—	—

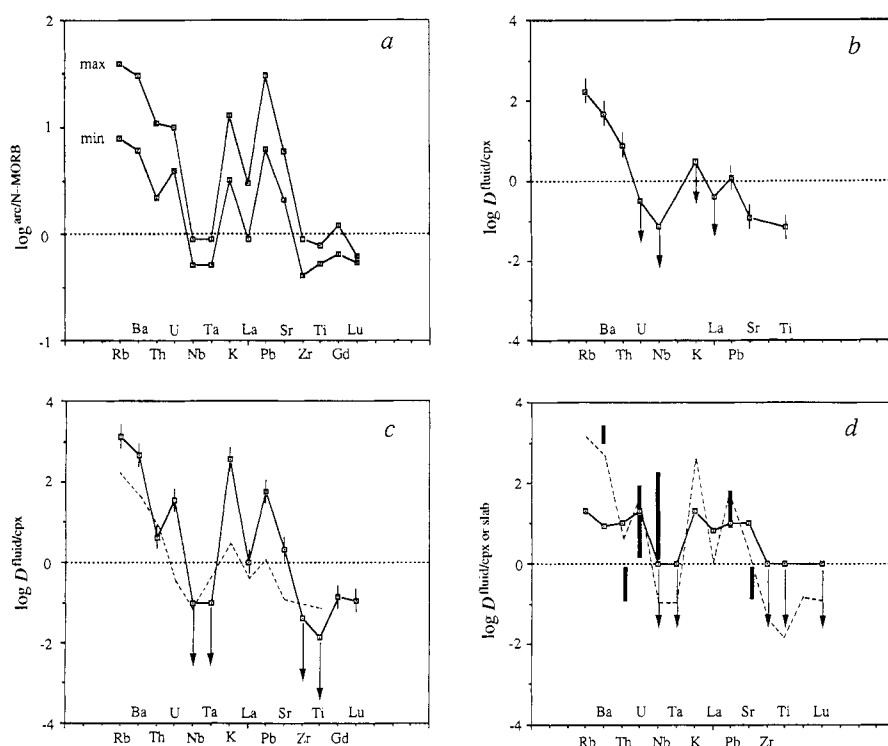
Note that $D^{\text{fluid/cpx}}$ (rightmost column) was calculated from $D^{\text{fluid/cpx}} = D^{\text{fluid/melt}} / D^{\text{cpx/melt}}$. $D^{\text{fluid/melt}}$ was measured in this work at 3 kbar and 1,040 °C. Analytical data are averages of 3–5 analyses. Relative standard deviations in per cent are given in parentheses. Normal experiments contained all trace elements doped into the glass phase at the beginning. * Indicates reversed experiments with trace elements initially doped into the hydrous solution only; Nb was added as layer of solid Nb₂O₅ at the glass/fluid interface. Owing to the longer run durations, the results of the normal experiments should be closer to equilibrium.

or a 5 m (Na,K)Cl solution containing 154 g NaCl and 175 g KCl per kg H₂O were added to the capsule. The synthetic andesitic glass used as starting material had a composition of 62.4 wt% SiO₂, 18.5 wt% Al₂O₃, 5.6 wt% MgO, 6.2 wt% CaO, 4.3 wt% Na₂O, 1.2 wt% K₂O and was doped with several trace elements, mostly at the 1,000 p.p.m. level. Experiments at 3 kbar were carried out in rapid-quench molybdenum alloy (TZM) vessels using argon as pressure medium. After the experiments, capsules were opened and the contents boiled for a few minutes with half-concentrated hydrochloric acid in order to dissolve material that precipitated from the fluid during the fast (1–2 s) quench¹⁰. Test experiments showed that during this acid treatment, only insignificantly small amounts of trace elements are leached from the glass phase. Both quenched fluid and glass phase were analysed by inductively coupled plasma atomic emission spectroscopy (ICP-AES). The results of these experiments are given in Table 1. Attainment of equilibrium was confirmed by true reversals, starting with the trace elements doped into the fluid. Experiments at 10–20 kbar were conducted in a piston-cylinder apparatus with a talc-pyrex cell. In these experiments, glasses were carefully analysed before and after the run by electron microprobe using very long total counting times; partition coefficients were calculated using mass balance. Due to analytical uncertainties, partition coefficients < 0.1 cannot be resolved in the experiments at 10–20 kbar (Table 2). But the

data in Table 2 suggest that, although there is certainly a gradual change of fluid/melt partition coefficients with pressure, the order of magnitude of $D^{\text{fluid/melt}}$ for each element does not fundamentally change between 3 and 20 kbar. Because the solubility of silica in hydrous fluids greatly increases with pressure, these data also suggest that dissolved silica does not have a major effect on the partition coefficients, at least over the range of pressures and temperatures studied. Therefore, in the following discussion the 3-kbar data will be used, as they can be measured with high precision.

Experimentally determined fluid/clinopyroxene partition coefficients for a chloride-free fluid are shown in Fig. 1b. A brief inspection of these data shows that they cannot explain the enrichment pattern in Fig. 1a. All elements that plot above the dashed line in Fig. 1a should be easily transported by the fluid. This is only possible if the fluid/clinopyroxene partition coefficient is significantly larger than unity, that is it should plot above the dashed line in Fig. 1b. Only Rb, Ba and Th fulfill this condition, while there is a clear mismatch between the predicted and observed enrichment pattern for U, K, Pb and Sr. Note that the data in Fig. 1b not only give too-low numerical values for many partition coefficients, but also the relative fractionations between several elements are incorrectly predicted, for example for Pb and Th.

FIG. 1 a, Trace element abundances in subduction-zone volcanics normalized to normal mid-ocean-ridge basalts (N-MORBs). Elements are ordered according to their compatibility during fluid-absent melting. Shown are minimum and maximum abundances for average samples from the New Britain, Kermadec, Marianas, Aleutian, and Vanuatu island arcs¹⁶. Note the pronounced negative spike at Nb and Ta, known as the "Nb-Ta anomaly". b, Partition coefficients of trace elements between a chloride-free hydrous fluid and clinopyroxene at 3 kbar and 1,040 °C. Analytical errors are smaller than the symbols shown; error bars give maximum uncertainties due to possible incomplete attainment of equilibrium, as derived from reversed experiments. Arrows indicate that the data given are only upper limits for the partition coefficients of these elements. c, Partition coefficients of trace elements between a 5 m (Na,K)Cl fluid and clinopyroxene at 3 kbar and 1,040 °C. Error bars and arrows as in b. Data for chloride-free hydrous fluids from b are shown for comparison (dashed curve). Note that a larger number of elements has been measured in the chloride-bearing experiments to check the consistency of these data with the observed arc enrichment pattern. d, Estimated bulk fluid/slab partition coefficients of McCulloch and Gamble¹⁶ (dots) and ranges of fluid/clinopyroxene partition coefficients measured by Brenan *et al.*¹² in recrystallization experiments at 900 °C and 20 kbar with a chloride-free hydrous fluid (black bars). Arrows as in b. Shown as a dashed curve are the fluid/clinopyroxene partition coefficients for a 5 m (Na,K)Cl fluid from c. In the McCulloch and Gamble pattern, the partition coefficients of Nb and Ta were assumed to be identical and data for La and Lu were estimated from



partition coefficients for Ce and Yb. Note that the data of Brenan *et al.* would require the stability of rutile in the subducted slab in order to explain the negative Nb-Ta anomaly in arc magmas.

Figure 1c gives the results of the partitioning experiments with 5 m (Na,K)Cl fluid. By the addition of chloride to the fluid, many partition coefficients were increased by orders of magnitude (for example, U, K, Pb) while others remained almost unaffected. Comparison with Fig. 1a shows that these data almost perfectly reproduce the enrichment pattern observed in subduction-zone magmas. The relative fractionations of almost all elements are correctly predicted. Moreover, the numerical values for the partition coefficients of the most mobile elements Ba, Rb, U, K and Pb are close to those predicted by theoretical modelling of the mass transport by fluids in a subduction zone. According to Hawkesworth *et al.*⁴, the fluid/mantle-wedge partition coefficients of these elements should approximately reach or exceed 100, in agreement with the experimental data.

Figure 1d compares the data of the present study with previous estimates of bulk fluid/slab or fluid/clinopyroxene partition coefficients. The fluid/slab partition coefficients estimated by McCulloch and Gamble¹⁶ (dots in Fig. 1d) are qualitatively similar to the results shown in Fig. 1c. Ranges of fluid/clinopyroxene partition coefficients measured by Brenan *et al.* in recrystallization experiments without chloride are shown as black bars. For some elements (for example, Ba), the results are similar to the present study, while there are discrepancies for other elements, such as Nb and Th. However, the two data sets may not be directly comparable, as the amounts of dissolved Si and Al in the fluid during the experiments may have been different. Brenan *et al.*¹² suggest that apart from chloride, the presence of these solutes may significantly affect partition coefficients. More recently, Ayers and Eggler¹⁸ suggested fluid/melt partition coefficients to be closer to unity at 1,250 °C and 15–20 kbar. The discrepancy between their results and those reported here may result from differences in experimental design.

One of the most interesting geochemical features of calc-alkaline magmas and of the entire continental crust is the so-called

"lead paradox"^{2,5,17}. Although Pb behaves less incompatibly than U and Th during formation of oceanic crust, the relative compatibilities are reversed for the formation of continental crust in subduction zones. Table 1 provides a simple explanation for this observation: in equilibrium with a silicate melt or with minerals such as clinopyroxene, Pb partitions much more strongly into a chloride-rich fluid than either U or Th. Metasomatism by such a fluid will therefore transport much more Pb than U or Th from the subducted slab or the mantle wedge into the zone of melting. It should be noted that the experiments in Table 1 were conducted without true buffering of oxygen fugacity; the oxygen fugacity probably was much higher than according to the Ni–NiO buffer, implying that the fluid/melt and fluid/clinopyroxene partition coefficients for U are probably overestimated. This means that

TABLE 2 Pressure dependence of partition coefficients between andesitic melt and 5 m (Na,K)Cl-fluid

Element	$D^{\text{fluid/melt}}$ 3 kbar 1,040 °C	$D^{\text{fluid/melt}}$ 10 kbar 1,100 °C	$D^{\text{fluid/melt}}$ 15 kbar 1,100 °C	$D^{\text{fluid/melt}}$ 20 kbar 1,100 °C
K	2.7	2.5	1.8	1.6
Rb	2.7	> 8	> 8	> 8
Sr	0.27	0.23	0.65	0.50
Ba	0.31	0.17	0.72	0.59
La	0.056	< 0.1	< 0.1	< 0.1
Gd	0.052	< 0.1	< 0.1	< 0.1
Lu	0.047	< 0.1	< 0.1	< 0.1
Ta	< 0.004	< 0.1	< 0.1	< 0.1

Experiments at 3 kbar: run duration 20 d; analysis of both phases by ICP-AES. Experiments at 10–20 kbar: run duration 24–66 h; analysis of the quenched glass by electron microprobe, calculation of partition coefficients by mass balance.

the fractionation of Pb and U during fluid transport is even larger than suggested by the data in Table 1.

In equilibrium with clinopyroxene and other minerals such as olivine and orthopyroxene, Sr partitions in favour of a hydrous chloridic fluid (Table 1). This provides a mechanism that allows contamination of the Sr isotopes in the zone of melting by Sr from the subducted oceanic slab, in agreement with the relatively radiogenic Sr-isotope composition of many calc-alkaline magmas^{4,6,7}. Another distinctive feature of such magmas are often radioactive disequilibria²² with strong enrichments of ²²⁶Ra over ²³⁰Th. These disequilibria may also be a result of fluid transport. Radium behaves geochemically very similar to Ba. From the data in Table 1, one would therefore expect that a hydrous chloridic fluid would strongly enrich ²²⁶Ra over ²³⁰Th, as is observed¹⁹.

From the foregoing discussion, it appears that the entire trace-element enrichment pattern in calc-alkaline magmas could be explained by metasomatism involving an alkali-chloride-rich fluid. This would imply that ultimately the trace element composition of the continental crust is largely a result of the complexing properties of chloride in supercritical aqueous solutions. Because the composition of the fluid released in the subducted slab depends on the composition of the sea water the oceanic crust had been reacting with, very old calc-alkaline rocks may preserve a record of the chemical evolution of sea water. The development of the typical modern trace element signature in subduction-zone vol-

canics may mark the point in geological history when sea water approached its modern composition, rather than reflecting a change in the style of global tectonics. □

Received 3 July 1995; accepted 1 February 1996.

1. Davies, J. H. & Stevenson, D. J. *J. geophys. Res.* **97**, 2037–2070 (1992).
2. Tatsumi, Y., Hamilton, D. L. & Nesbitt, R. W. *J. Volcan. geotherm. Res.* **29**, 293–309 (1986).
3. Maury, R. C., Defant, M. J. & Joron, J. L. *Nature* **360**, 661–663 (1992).
4. Hawkesworth, C. J., Gallagher, K., Hergt, J. M. & McDermott, F. *Lithos* **33**, 169–188 (1994).
5. Miller, D. M., Goldstein, S. L. & Langmuir, C. H. *Nature* **368**, 514–520 (1994).
6. Perfit, M. R., Gust, D. A., Bence, A. E., Arculus, R. J. & Taylor, S. R. *Chem. Geol.* **30**, 227–256 (1980).
7. Arculus, R. J. *Lithos* **33**, 189–208 (1994).
8. Philippot, P. *Chem. Geol.* **108**, 93–112 (1993).
9. Brenan, J. M. & Watson, E. B. *Earth planet. Sci. Lett.* **107**, 672–688 (1991).
10. Keppeler, H. & Wyllie, P. J. *Contr. Miner. Petrol.* **109**, 139–150 (1991).
11. Hunnee, J. E. *Inorganic Chemistry* (Harper & Row, New York, 1983).
12. Brenan, J. M., Shaw, H. F., Ryerson, F. J. & Phinney, D. L. *Geochim. cosmochim. Acta* **59**, 3331–3350 (1995).
13. Johnson, K. T. M. & Kinzler, R. J. *Eos* (abstr.) **70**, 1388 (1989).
14. Chausson, M. & Libourel, G. *Geochim. cosmochim. Acta* **57**, 5053–5062 (1993).
15. Carroll, M. R. & Wyllie, P. J. *Am. Mineral.* **75**, 345–357 (1990).
16. McCulloch, M. T. & Gamble, J. A. *Earth planet. Sci. Lett.* **102**, 358–374 (1991).
17. Hofmann, A. E. *Earth planet. Sci. Lett.* **90**, 297–314 (1988).
18. Ayers, J. C. & Eggler, D. H. *Geochim. cosmochim. Acta* **59**, 4237–4246 (1995).
19. Condomines, M., Hemond, C. & Allègre, C. J. *Earth planet. Sci. Lett.* **90**, 243–262 (1988).
20. Hart, S. R. & Dunn, T. *Contr. Miner. Petrol.* **113**, 1–8 (1993).
21. Irving, A. J. *Geochim. cosmochim. Acta* **42**, 743–770 (1978).
22. Green, T. H. *Chem. Geol.* **117**, 1–36 (1994).

ACKNOWLEDGEMENTS. I thank A. Dietel for carrying out all ICP-analyses reported here and D. Krauss for his assistance with the electron microprobe. A review by T. Plank improved the manuscript. This work was supported by the German Science Foundation (DFG).

The rock–paper–scissors game and the evolution of alternative male strategies

B. Sinervo & C. M. Lively

Department of Biology and Center for the Integrative Study of Animal Behavior, Indiana University, Bloomington, Indiana 47405, USA

MANY species exhibit colour polymorphisms associated with alternative male reproductive strategies, including territorial males and ‘sneaker males’ that behave and look like females^{1–3}. The prevalence of multiple morphs is a challenge to evolutionary theory because a single strategy should prevail unless morphs have exactly equal fitness^{4,5} or a fitness advantage when rare^{6,7}. We report here the application of an evolutionary stable strategy model to a three-morph mating system in the side-blotched lizard. Using parameter estimates from field data, the model predicted oscillations in morph frequency, and the frequencies of the three male morphs were found to oscillate over a six-year period in the field. The fitnesses of each morph relative to other morphs were non-transitive in that each morph could invade another morph when rare, but was itself invadable by another morph when common. Concordance between frequency-dependent selection and the among-year changes in morph fitnesses suggest that male interactions drive a dynamic ‘rock–paper–scissors’ game⁷.

From 1990–95, we studied territory use and patterns of sexual selection on male side-blotched lizards (*Uta stansburiana*), a small territorial iguanid lizard (3–10 g) that matures in one year in the inner Coast Range of California (Merced County). Territory defence by males is dependent on a throat-colour polymorphism that develops as males mature (Fig. 1). Males with orange throats are very aggressive and defend large territories. Males with dark blue throats are less aggressive and defend smaller territories. Males with prominent yellow stripes on their throats are ‘sneakers’ and do not defend territories (receptive females also have yellow-striped throats). Aggression in orange males is linked to higher levels of testosterone⁸, as is the case for other vertebrates^{3,9–12}.

Male fitness was estimated by the number of females that are exclusively on his home range (monopolized females), plus his share of females (shared females) that overlap with the home range of other males (Fig. 3a–e). Significant changes in morph frequency among years were consistent with an evolutionary response to sexual selection (Fig. 3), particularly in light of the heritability ($h^2 = 0.96$) for throat colour in nature (Fig. 2). Fitness regression coefficients¹³ describe the impact that neighbouring morphs have on male fitness. ‘Local’ frequency-dependent sexual selection among morphs could drive frequency fluctuations among years (Fig. 3). The population cycled from high frequency of blue (1991), to high frequency of orange (1992) to high frequency of yellow (1993–94) and returned to high frequency of blue (1995).

Orange males increased from 1991 to 1992, owing to their significantly higher monopoly on females in 1991 (Fig. 3a). Moreover, the decline in blue males from 1991 to 1994 arose from losses to adjacent orange males. In 1992, blue males lost a monopoly on 0.11 ($P < 0.05$) females for every orange neighbour (0.59 oranges on average). Conversely, in 1993 orange males gained a monopoly on 1.10 females ($P < 0.01$) for each blue neighbour (0.09 blues on average). After peaking in 1992, orange males declined in frequency from 1992 to 1995, and their decline was consistent with a drop in monopolization observed for orange males (Fig. 3a, b).

Although no morph had a clear fitness advantage in 1992 or 1993, the decline in orange males was coincident with an increase in the frequency of yellow males (Fig. 3c, d). Losses of orange males were directly related to the number of neighbouring yellows. In 1992, each orange neighbour increased a yellow male’s monopoly by 0.23 females ($P < 0.0001$) and a yellow male’s share by 0.54 females ($P < 0.0001$, 0.46 oranges on average). Conversely, in 1995 orange males lost a share of 0.55 females ($P < 0.01$) to each yellow neighbour (2.25 yellows on average).

Yellow males declined in frequency from 1993 to 1995, and their decline was due in part to lower monopolization of female home ranges by yellow males compared to blue or orange males (Fig. 3d). In addition, each blue male obtained 0.25 shared females ($P < 0.01$) for each yellow neighbour (0.63 yellows on average). In 1995, the profile of selection returns to the pattern observed in 1991, when orange males have high rates of female monopoly (compare Fig. 3e with a), and orange males should

FIG. 1 Colour polymorphisms of male side-blotched lizards. Males with orange throats (lower left panel) and/or sides (male with orange flank on the left, top panel) are 'ultra dominant' to males with blue on their throat (lower centre panel, and male on the right, top panel). Males with yellow throats (lower right panel) resemble females in morphology in that females have yellow throats when receptive. Whereas males with orange, blue and yellow throats represent discrete and unambiguous colour scores (orange, 2; blue, 1; yellow, 0) a small fraction of 'hybrid' blue males mature with dark blue throats and very thin orange stripes. These males were categorized as blue based on behaviour, and given a colour score of 1.5 for the heritability analyses. Likewise a small fraction of yellow-throated males mature with yellow throats with very pale blue stripes (not dark blue) and these males were categorized as yellow and given a colour score of 0.5 for the heritability analyses.

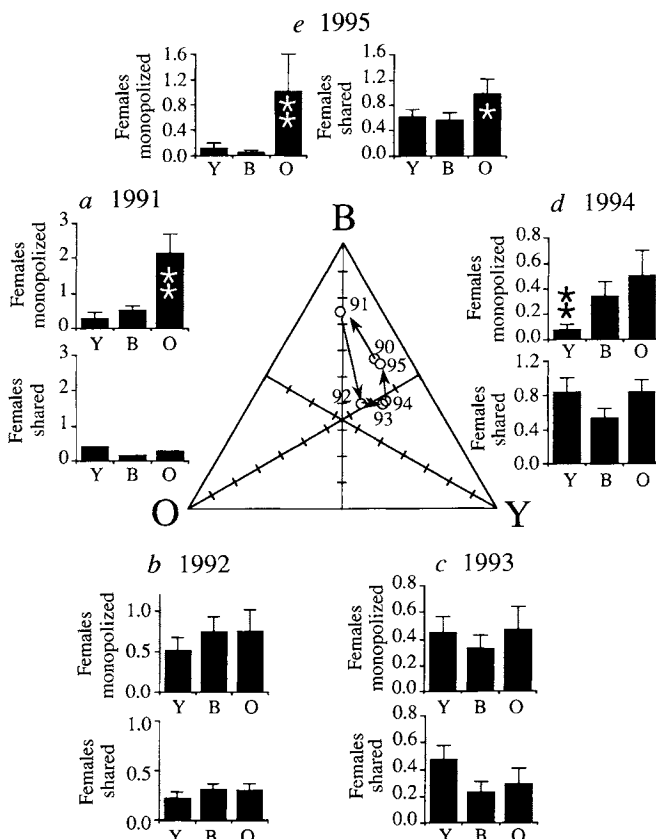
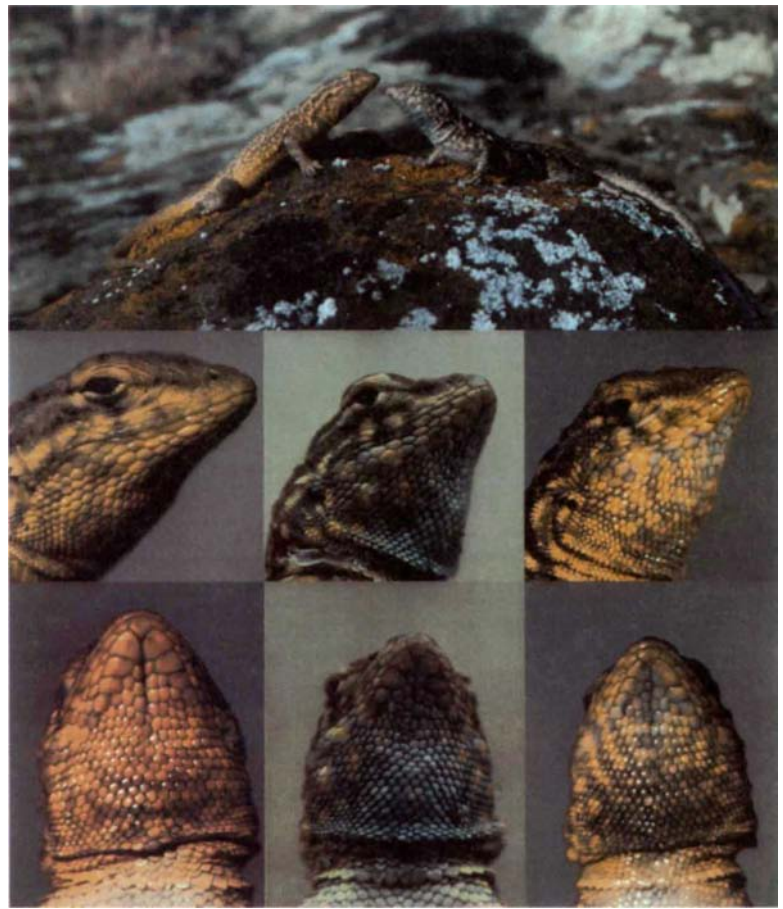
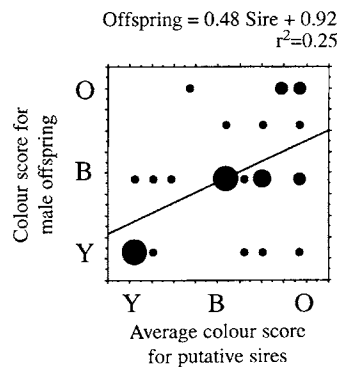


FIG. 2 During 1990–95, significant among-year changes in the frequency of adult male colour polymorphisms (Fig. 1) on a 250-m-long sandstone outcropping (central triangle, Chi-square, $P < 0.01$) were consistent with significant variation in sexual selection (histograms a–e, number of females on a male's territory, $\bar{x} \pm \text{s.e.}$), and morph fitness (see text). Frequency of morphs are plotted for each year as follows: 0–100% blue from base to apex, 0–100% orange from right side to left vertex, and 0–100% yellow from left side to right vertex.

METHODS. A male monopolized a female if he was the only male observed on a female's home range. We partitioned females with many males on her home range among males (a mean which differs from 2 means is denoted by **, and * denotes a mean that differs from a single mean; Fisher's LSD post-hoc test, $P < 0.05$). Survival which is based on comprehensive monthly recaptures¹⁵, reproductive success^{15–17}, and territorial behaviour (discriminated with a dorsal paint-mark)^{14,15} of all uniquely marked adults was followed. Home ranges were mapped^{14,15} from daily visual censuses ($N_{\text{males}} = 62, 127, 105, 116, 74$; $N_{\text{females}} = 133, 189, 178, 140, 106$. $N_{\text{sightings}} = 518, 1,846, 1,791, 3,466, 1,664$ for 1991–95, respectively). Only females that survived to produce eggs were considered. Eggs were collected from all females and offspring were released upon hatching^{15–17}. Survival of uniquely marked offspring was censused at maturity the next year¹⁴.

FIG. 3 Heritability ($h^2 = 0.96$) of male throat colouration for animals in nature. Maternity of offspring on the site was known with certainty¹⁸. Eggs were obtained from uniquely marked females and their uniquely marked offspring were released randomly with respect to sibship back on site when they hatch^{14,18}. Throat colour of surviving male offspring was scored at maturity (Fig. 1). We assumed that a monopolizing male is the only sire of a dam's offspring. In cases of shared females, probability of paternity was assumed to be proportional to a putative sire's access to the dam. Regression of offspring and 'sire's' throat score yields $h^2 = 0.96$. Quantitative genetic analysis assumes many autosomal loci¹⁹, however, one or two loci may govern colour.



increase in 1996. However, given that each orange male lost a share of 0.55 females to each yellow neighbour (see above), we also predict that yellows should 'hitchhike a fitness ride' from oranges and increase in frequency relative to blues, as in 1991.

The pattern of frequency-dependent sexual selection among years leads to a key prediction: when a morph reaches a low frequency, it should produce the most mature offspring in the next year (for example, at the 'vertices' of the cycle, Fig. 3). Indeed, analysis of fitness, as measured by survival of offspring from putative sires, showed a significant morph $\times$ year interaction ($P < 0.0001$, morph fitness, $W_i = W_{i:\text{rare}}/W_{i:\text{common}}$). In 1991, orange males produced the most offspring that survived to the next year ($W_o = 4.73$, $W_b = 1.00$, $W_y = 1.61$), yellow males produced the most in 1992 ($W_o = 1.28$, $W_b = 1.00$, $W_y = 1.45$), and blue males produced the most in 1994 ($W_o = 0.90$, $W_b = 1.26$, $W_y = 1.00$). Differences in fitness of putative sires were consistent with sexual selection (above) and were not due to survival differences of offspring to maturity ($P > 0.36$, $N = 1,202$).

We estimated the frequency-dependent 'fitness payoffs' of a

rare morph competing against a common morph⁷ from the 'local' frequency-dependent interactions among males (Fig. 4). The fitness of a rare i^{th} strategy against a common (j^{th}) strategy was estimated as:

$$W_{ij} = [M_i + N_i \times G_{Mij}] + [S_j + N_j \times G_{Sij}] \quad (1)$$

where fitness, W_{ij} , is the sum of two linear equations. The linear equation in the left brackets is fitness acquired through monopoly of female territories, M_i , adjusted by G_{Mij} which is the slope of the line (partial regression coefficients, above) describing gains and losses for the i^{th} strategy for each neighbour of the j^{th} strategy. The number of neighbours, N_i , is a phenotypic trait of the i^{th} strategy. N_i describes the total number of males that the i^{th} strategy is playing against ($N_y = 5.03$, $N_b = 2.35$ and $N_o = 2.95$ from 1991–95). This assumes gains and losses are linearly related to the number of neighbouring males. Likewise, fitness from shared access to females, S_j , is also adjusted by G_{Sij} and N_j . Equation (1) can be written in a compact matrix form where, $\mathbf{W}$, is a pay-off matrix (Fig. 4a) that describes the relative fitness of each rare morph in competition with a common morph (elements $i = 1, 2$, and 3 of $\mathbf{W}$ correspond to strategies y, b, and o):

$$\mathbf{W} = \mathbf{M} + \mathbf{S} + \mathbf{N} \cdot (\mathbf{G}_M + \mathbf{G}_S) \quad (2)$$

When the rare and common morph are the same strategy, they have equal fitness (diagonal elements of $\mathbf{W}$ are equal to 1.0, Fig. 4a). To be an evolutionary stable strategy (ESS), a morph must have higher fitness than the other morphs when it is both rare and common. It is clear that no morph is an ESS (Fig. 4). When generations are discrete, the population may be expected to oscillate around an 'attractor'. Assuming $h^2 = 1$, the frequency of the i^{th} morph in the next generation, $p_{i,t+1}$, can be written as⁶:

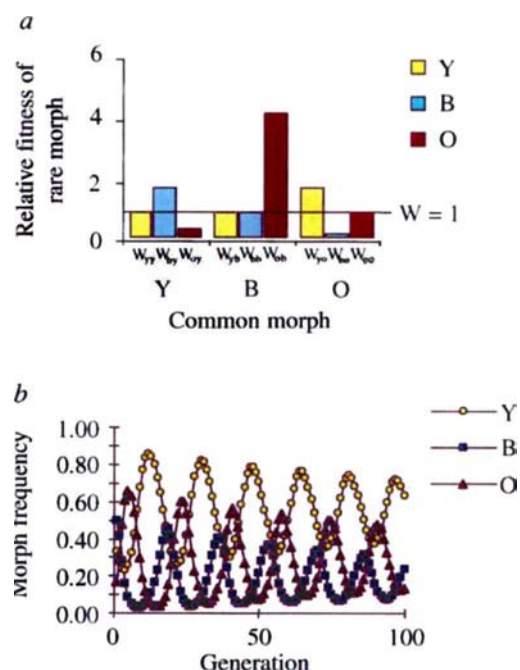
$$p_{i,t+1} = p_{i,t} \left(\frac{W_{i,t}}{\bar{W}} \right) \quad (3)$$

where $W_{i,t} = p_{1,t}W_{i1} + p_{2,t}W_{i2} + p_{3,t}W_{i3}$ (where W_{i1} , W_{i2} , and W_{i3} are elements of $\mathbf{W}$, equation (2)), and $\bar{W} = \sum p_{i,t} \times W_{i,t}$. Iterations of the model using parameter estimates from field data, give oscillations qualitatively similar (Fig. 4b) to those observed in the field (Fig. 2). The model also shows that no morph is an ESS, which

FIG. 4 a, The pay-off matrix of a rare yellow, blue or orange strategy playing against each strategy when common. The strategy (yellow) is an ESS if it cannot be invaded when common by either a rare orange or a rare blue strategy: $W_{yy} > W_{oy}$, and $W_{yy} > W_{by}$. Similarly, blue is an ESS if $W_{bb} > W_{ob}$, and $W_{bb} > W_{yb}$; and orange is an ESS if $W_{oo} > W_{yo}$, and $W_{oo} > W_{bo}$. It is clear that every morph is invadable by one other morph and thus no morph is an ESS. Analysis of the pay-off matrix, $\mathbf{W}$, was partitioned according to access to monopolized and shared females (equation (2) and Fig. 2). We observed significant frequency-dependence in every year (see text). Partial regression coefficients (adjusted to relative fitness) describing gains and losses are given by:

$$N(\mathbf{G}_M + \mathbf{G}_S) = \begin{bmatrix} 5.03 & 0 & 0 \\ 0 & 2.35 & 0 \\ 0 & 0 & 2.95 \end{bmatrix} \times \left(\begin{bmatrix} 0 & 0 & +0.15 \\ 0 & 0 & -0.07 \\ 0 & 1.24 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & +0.35 \\ +0.27 & 0 & 0 \\ -0.93 & 0 & 0 \end{bmatrix} \right)$$

For example, a rare yellow would gain 0.15 females by monopoly and 0.35 females by sharing when playing against common orange males. A rare yellow's gains in relative fitness are equal to $5.03 \times (0.15 + 0.35)$, given 5.03 orange neighbours for the rare yellow. Fitness gains and losses between rare and common neighbours, $N(\mathbf{G}_M + \mathbf{G}_S)$, are used to adjust relative fitness of each morph (for example, female monopoly and shares, Fig. 2). b, The empirically derived pay-off matrix, $\mathbf{W}$, was used in a simple quantitative genetic model (equation (3)). The model predicts cycles in morph frequency with a period of 12 years. Although the magnitude of oscillations appears to damp over time, any random perturbation of frequency will renew the strength of the oscillations.



leads to dynamic oscillations in frequency (Fig. 4b). Oscillations in morph frequency (not shown) are surprisingly robust in an age-structured model that includes small, but significant differences in yearly adult survival between morphs ($l_y = 0.167$, ($N = 209$), $l_b = 0.136$ ($N = 199$), and $l_o = 0.048$ ($N = 84$), Chi-square: l_o versus l_y , $P < 0.01$, l_o versus l_b , $P = 0.05$, l_y versus l_b , n.s.):

$$p_{i,t+1} = [l_i \times p_{i,t}] + (1 - l_{\text{adult}}) \times p_{i,t} \left(\frac{W_{i,t}}{\bar{W}} \right) \quad (4)$$

where, $p_{i,t+1}$ is determined by the probability of adult morph survival to the next year (l_i) and relative fitness of each morph (equation (3)) is adjusted by the proportion of new recruits in the

next year, which is approximated by $(1 - l_{\text{adult}}) = 1 - \sum p_{i,t} \times l_i$, or one minus average adult survival.

We have described the first biological example of a cyclical 'Rock-paper-scissors' game⁷. As in the game where paper beats rock, scissors beat paper, and rock beats scissors, the wide-ranging 'ultradominant' strategy of orange males is defeated by the 'sneaker' strategy of yellow males, which is in turn defeated by the mate-guarding strategy of blue males; the orange strategy defeats the blue strategy to complete the dynamic cycle. Frequency-dependent selection maintains substantial genetic variation in alternative male strategies, while at the same time prohibiting a stable equilibrium in morph frequency. □

Received 31 October 1995; accepted 18 January 1996.

- Gross, M. R. *Phil. Trans. R. Soc. B* **332**, 59–66 (1991).
- Moore, M. C. & Thompson, C. W. in *Progress in Clinical and Biological Research* (eds Eppele, A., Scanes, C. G. & Stetson, M. H.) 685–690 (1990).
- Brantley, R. K., Wingfield, J. C. & Bass, A. H. *Horm. Behav.* **27**, 332–347 (1993).
- Ryan, M. J., Pease, C. M. & Morris, M. R. *Am. Nat.* **139**, 21–31 (1992).
- Shuster, S. M. & Wade, M. J. *Nature* **350**, 608–610 (1991).
- Charnov, E. L. *Life History Invariants: Some Explorations of Symmetry in Evolutionary Ecology* 1–167 (Oxford Univ. Press, New York, 1993).
- Maynard-Smith, J. *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, 1982).
- Sinervo, B., Miles, D. B., DeNardo, D., Frankino, T. & Klukowski, M. *Proc. R. Soc. Lond.* (submitted).
- Kindler, P. M., Phillip, D. P., Gross, M. R. & Bahr, J. M. *Gen. comp. Endocrinol.* **75**, 446–453 (1989).

- Rand, M. S. *Gen. comp. Endocrinol.* **88**, 461–468 (1992).
- Cardwell, J. R. & Liley, N. R. *Gen. comp. Endocrinol.* **81**, 7–20 (1991).
- Hews, D. K., Knapp, R. & Moore, M. C. *Horm. Behav.* **28**, 96–115 (1994).
- Lande, R. & Arnold, S. J. *Evolution* **37**, 1210–1226 (1983).
- Sinervo, B., Doughty, P., Huey, R. B. & Zamudio, K. *Science* **258**, 1927–1930 (1992).
- Sinervo, B. & DeNardo, D. F. *Evolution* (in the press).
- DeNardo, D. F. & Sinervo, B. *Horm. Behav.* **28**, 53–65 (1994).
- DeNardo, D. F. & Sinervo, B. *Horm. Behav.* **28**, 273–287 (1994).
- Sinervo, B. & Doughty, P. *Evolution* (in the press).
- Falconer, D. S. *Introduction to Quantitative Genetics* 1–340 (Longman, New York, 1981).

ACKNOWLEDGEMENTS. We thank P. Stadler and R. Schrimp for permission to work on their land. This research was supported by NSF grants to B.S. and C.M.L. Video segments of morph head-bob patterns can be obtained in Quick Time format through the World Wide Web (<http://sunflower.bio.indiana.edu/~bsinervo>).

A role for melanin-concentrating hormone in the central regulation of feeding behaviour

Daqing Qu, David S. Ludwig*, Steen Gammeltoft†, Megan Piper, Mary Ann Pelleymounter‡, Mary Jane Cullen‡, Wendy Foulds Mathes§, Jeanne Przypek§, Robin Kanarek§ & Eleftheria Maratos-Flier||

Elliott P. Joslin Research Laboratory, Joslin Diabetes Center, and Department of Medicine, Brigham and Women's Hospital, Boston, Massachusetts 02215, USA

* Department of Medicine, Division of Endocrinology, Beth Israel Hospital, Boston, Massachusetts 02215, USA

‡ Department of Neurobiology, Amgen Inc., 1840 DeHavilland Drive, Thousand Oaks, California 91320, USA

§ Department of Psychology, Tufts University, Medford, Massachusetts 02155, USA

The hypothalamus plays a central role in the integrated regulation of energy homeostasis and body weight, and a number of hypothalamic neuropeptides, such as neuropeptide Y (ref. 1), galanin², CRH (ref. 3), and GLP-1 (ref. 4), have been implicated in the mediation of these effects. To discover new hypothalamic peptides involved in the regulation of body weight, we used differential display polymerase chain reaction⁵ to identify messenger RNAs that are differentially expressed in the hypothalamus of *ob/ob* compared with *ob/ob* C57Bl/6J mice. We show here that one mRNA that is overexpressed in the hypothalamus of *ob/ob* mice encodes the neuropeptide melanin-concentrating hormone (MCH). Fasting further increased expression of MCH mRNA in both normal and obese animals. Neurons containing MCH are located in the zona incerta and in the lateral hypo-

thalamus. These areas are involved in regulation of ingestive behaviour, but the role of MCH in mammalian physiology is unknown. To determine whether MCH is involved in the regulation of feeding, we injected MCH into the lateral ventricles of rats and found that their food consumption increased. These findings suggest that MCH participates in the hypothalamic regulation of body weight.

Using differential display polymerase chain reaction (PCR) with 180 primer pairs, we screened 30% of expressed mRNAs. Fifty-two DNA bands appeared to be differentially expressed. Of 35 bands further evaluated using northern blot analysis with riboprobes, no signal could be detected for 9 bands, and no difference in expression was observed in 20 bands. Thus, of about 9,000 complementary DNAs screened, differences in expression were confirmed for only six bands (about 0.7%). Of these, two had matches in Gene Bank: one was MCH (Fig. 1) and the other was the mouse oncogene, *fau*. A third band had homology to a DNA-binding factor, and three additional bands that were differentially expressed had no known homology. Although the difference in MCH expression on differential displays appeared to be absolute, that is, no signal was detected in the *ob/ob* mice versus an obvious signal in *ob/ob* mice, assessment of MCH expression using a ribonuclease protection assay showed that the difference between fed *ob/ob* and *ob/ob* mice was a 50–80% increase in the *ob/ob* animal (Fig. 1b).

To evaluate further the expression of MCH in lean versus obese mice and to evaluate the possibility that MCH expression was affected by nutritional status, C57Bl/6J mice +/+ , C57Bl/6J *ob/ob* heterozygotes and C57Bl/6J *ob/ob* animals were compared both in the fed state and after 24 hours of fasting. In these experiments we also probed for neuropeptide Y (NPY) mRNA because it is known that NPY expression increases with fasting and is also increased in *ob/ob* mice. Figure 2 shows the results of hypothalamic mRNA blots in fed and fasted mice probed for MCH or NPY. In the fed state, wild-type control mice express low levels of MCH, whereas *ob/ob* animals and *ob/ob* animals express higher levels of MCH. Fasting led to an increase in MCH expression in all three groups of animals.

Figure 3 shows the quantitative data derived from several experiments. Figure 3a and b compares relative expression of MCH in the fed and fasted state in control, heterozygous and

† Present address: Department of Clinical Chemistry, Glostrup Hospital, University of Copenhagen Medical School, DK-2600 Glostrup, Denmark.

|| To whom correspondence should be addressed.

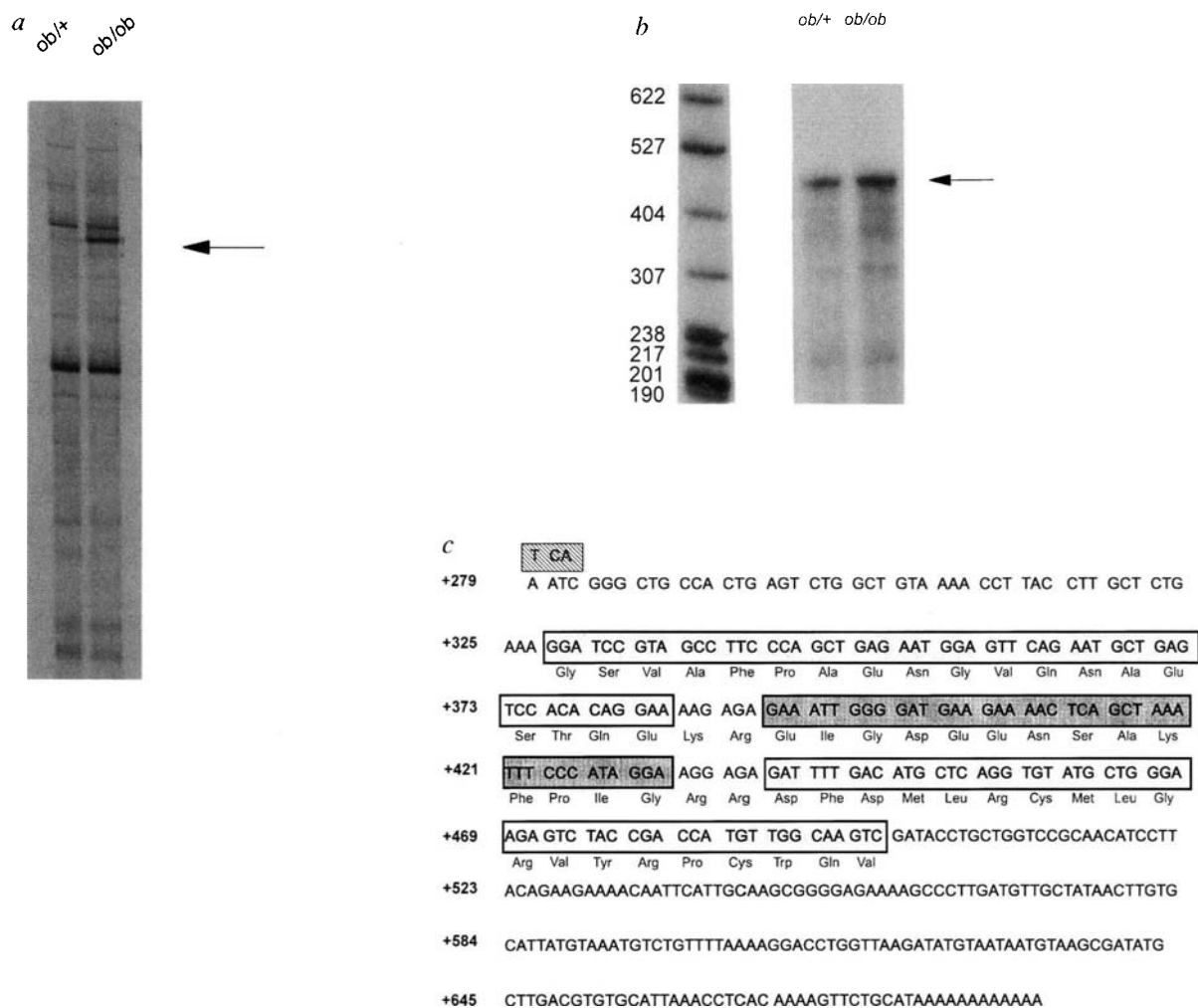


FIG. 1 **a**, PCR display on hypothalamic RNA from fed *ob/+* and *ob/ob* mice was performed using 5'-TTTTTTTTTTTAC-3' as the anchored primer and 5'-AATCGGGCTG-3' as an arbitrary primer. Candidate gene is indicated by an arrow. **b**, Ribonuclease protection assay, using the amplified band inserted into pCR-I plasmid, was performed to confirm differential expression of MCH in *ob/+* vs *ob/ob* mice. Fed *ob/ob* mice expressed ~40% more MCH mRNA than did fed *ob/+*. **c**, Sequence of candidate band inserted into pCR-I plasmid. The shaded areas represent coding sequences for N-EG (bases 328–384), N-EI (bases 391–432) and MCH (bases 439–495). The cross-hatched area over the first three bases shows the three mismatched bases. Actual bases of MCH gene are shown in the cross-hatched rectangle. The mismatch in the band amplified from RT-PCR display matches the sequence of the arbitrary primer used.

METHODS. Male *ob/ob* and *ob/+* heterozygotes C57Bl/6J mice were obtained at 7 weeks of age from Jackson laboratories. Mice were housed for at least 4 days after arrival, to allow them to recover from shipping. Fed mice were killed in the morning after being anaesthetized with an i.p. injection of 200 mg kg⁻¹ sodium amytal. Mice were decapitated, the brain was removed, and the hypothalamus identified, excised, homogenized and extracted immediately in RNazol (Cinna/Biotex Laboratories). Ten hypothalami, weighing ~100 mg were collected. Aliquots of total RNA were treated with DNase I (Boehringer-Mannheim) to remove any traces of DNA. RNA was divided into nine pools, and a cDNA was synthesized using M-MLV reverse transcriptase (Superscript RNAase H-Reverse Transcriptase; Gibco BRL), and one of nine anchored primers (see below). The cDNA generated was used in PCR display. Twelve possible downstream primers with the sequence T₁₂VN (where V may be A, C or G and N may be any nucleotide) and termed anchored primers, were used in conjunction with ~50

upstream random primers, designated arbitrary primers, for PCR display. The arbitrary upstream primers did not contain more than 50% GC and had no internal homology. Using these 600 primer pairs, it was possible to assess expression of about 30,000 mRNAs (each primer pair yielded ~50 cDNA bands). We used 180 primers pairs consisting of one of nine anchored primers (T₁₂VV where V is A, C or G) and 20 arbitrary primers to screen mRNA expression in hypothalamus of obese versus non-obese animals. The cDNA generated from the reverse transcriptase reaction was amplified using AmpliTaq DNA polymerase (Perkin Elmer). Twenty-five cycles of PCR reactions were performed; cycles consisted of 30 s at 94 °C, 2 min at 40 °C and 30 s at 72 °C. Reactions were in the presence of ³⁵S-dATP (NEN, Boston), and products were separated on sequencing gels under denaturing conditions. Dried gels were exposed to Kodak X-OMAT AR film (Eastman Kodak) for 24 to 48 h. After development, DNA fragments from *ob/ob* and *ob/+* hypothalamus were compared. Bands unique to either *ob/ob* or *ob/+* hypothalamus were excised from the dried gel, extracted by boiling in 100 µl of TE buffer, and precipitated with ethanol in the presence of mussel glycogen (Boehringer-Mannheim). DNA was further amplified using the original set of primers used to generate the particular band, using the same thermal cycling conditions. Reaction products were run out on a 1% agarose gel and stained with ethidium bromide. Bands were excised from the gel, eluted, and ligated into the PCR-I plasmid (Invitrogen) and were sequenced using an ABI-373 automated sequencer and M13R and -21M13F as primers, to determine both sequence and orientation. Depending on orientation the Sp6 or T7 promoter was used to generate a riboprobe. Altered expression of bands was confirmed using a ribonuclease protection assay²⁹ or northern blot analysis with riboprobes.

homozygous animals. MCH mRNA is overexpressed in obese animals in both the fed and fasted states. In the fed state MCH expression is double in *ob/ob* mice compared to the wild-type *+/+* mice. A similar twofold elevation of MCH mRNA was seen in animals evaluated in the fasted state. Expression of MCH mRNA in the heterozygote appeared to be intermediate between control and homozygote *ob/ob* mice. Figure 3c and d evaluates expression of NPY mRNA levels in the same experiments. NPY levels in *ob/ob* mice were doubled over control animals in both the fed and fasted state; heterozygotes expressed the same amount of NPY mRNA as controls. Figure 3e and f evaluates the effect of fasting on MCH and NPY mRNA levels. In control wild-type animals, fasting increased MCH mRNA levels almost fourfold (range two-

to fivefold), whereas the increase with fasting in *ob/ob* animals was somewhat less than threefold. These changes were of greater magnitude than the changes seen in NPY mRNA levels, which increased slightly less than twofold with fasting. Extra-hypothalamic expression of MCH has been reported^{6,7}. Using northern blots loaded with 30 µg RNA and a riboprobe, we detected the MCH signal only in the hypothalamus (data not shown).

To evaluate the hypothesis that MCH is involved in the regulation of feeding, we administered MCH into the lateral ventricle of rats through preimplanted indwelling catheters (ICV). Administration of 5 µg MCH at the beginning of the feeding cycle roughly doubled caloric consumption (Table 1) above that seen in animals injected with carrier only. This effect was rapid and persisted over

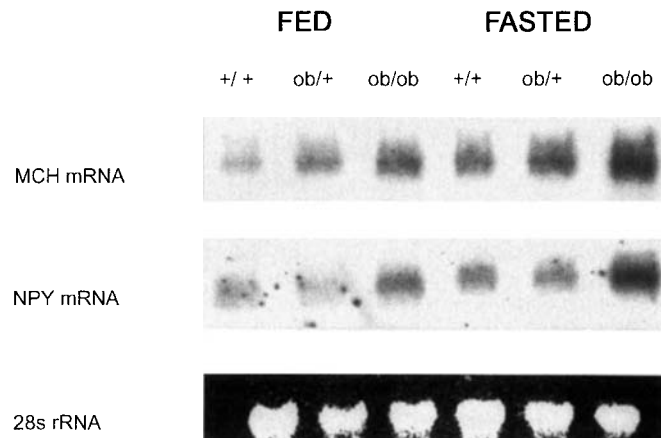


FIG. 2 Northern blot probed with riboprobes toward MCH or NPY. NPY was used as a control neuropeptide as NPY levels are increased in obese animals and regulated by fasting³⁰. Lower lanes show ethidium bromide staining of 28S ribosomal RNA before transfer. Both MCH and NPY mRNA levels were lowest in fed *+/+* animals. mRNA expression was higher in *ob/ob* animals and increased with fasting in all sets of animals.

METHODS. RNA was collected from fed and 24-h-fasted *+/+*, *ob/+* or *ob/ob* mice. Four mice were used in each group. RNA was subjected to agarose gel electrophoresis and transferred to a nylon membrane for probing. Northern blots containing 20 µg per lane of hypothalamic RNA from *ob/ob* or *ob/+* or *+/+* animals were analysed using a riboprobe generated from the pCR-I plasmid containing the 450-base fragment of MCH. Blots were prehybridized for 4 h at 58 °C; the probe was then added and filters were hybridized overnight at 58 °C, then washed with 2× SSC, 0.1% SDS at room temperature for 10 min followed by a wash with 0.1× SSC, 0.1% SDS at 60 °C for 10 min. This sequence of washes was repeated a total of three times. Northern blots were air-dried and exposed to either Kodak X-OMAT film or analysed using the Molecular Dynamics PhosphorImager.

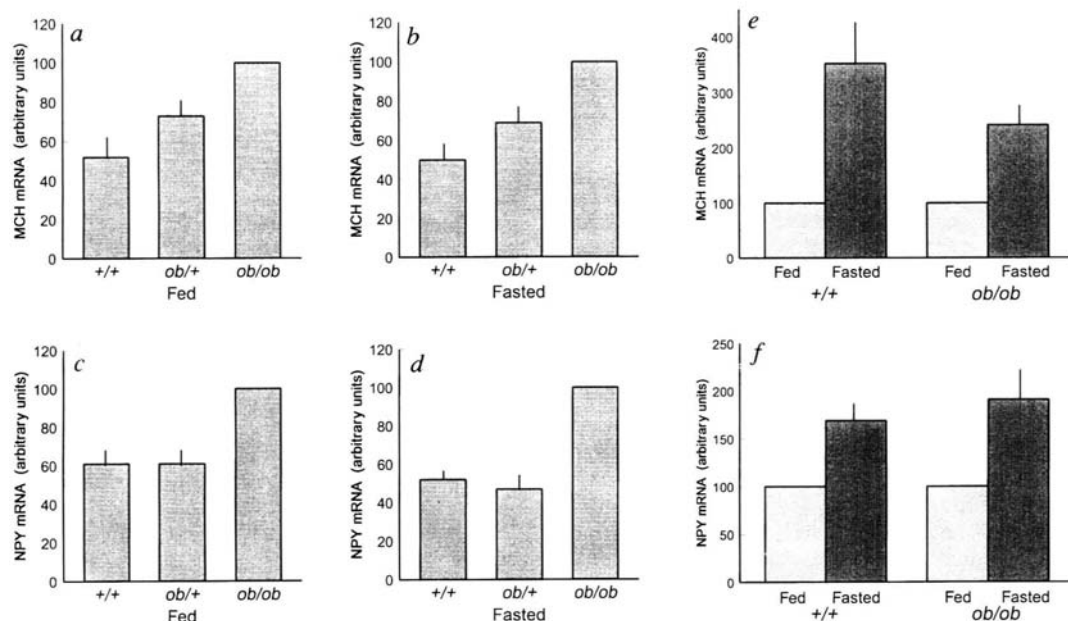


FIG. 3 Quantitative data obtained from multiple northern blots. In each panel the bar shows mean values and the error bar shows the standard error. a, Levels of MCH mRNA in fed wild-type, *ob/+* and *ob/ob* animals are compared; data from 4 blots. Each blot was scanned using the Molecular Dynamics PhosphorImager and in each experiment the intensity seen with *ob/ob* type animals set at 100%. The difference between control and *ob/ob* animals was analysed by paired t-tests and was statistically significant ($P = 0.032$). Although MCH levels in heterozygote mice tended to fall between those seen in wild-type and *ob/ob* mice the difference was not statistically significant ($P = 0.209$). In fasted animals the mean in MCH mRNA expression between wild-type and *ob/ob* animals was more than twofold (four separate experiments) that seen in b; the difference between wild-type and homozygote animals was highly significant ($P = 0.0053$). In

the fasted state, the difference in MCH mRNA expression in *ob/+* animals, which was intermediate between wild-type and *ob/ob* animals was significant ($P = 0.0113$). The increased expression of MCH mRNA in *ob/ob* animals was greater than the increase in the expression of NPY in the same animals (c, d). NPY experiments were only performed twice, as these data confirmed findings in the literature. e, Rise in MCH levels seen with fasting in both control (six experiments) and *ob/ob* animals (4 experiments). In each case the levels in fed animals are set at 100%. Fasting normal animals for 24 h led to a more than threefold increase in MCH mRNA expression (range two- to fivefold $P = 0.0061$). In *ob/ob* animals MCH mRNA levels more than doubled ($P = 0.018$). The magnitude of these changes was greater than those in NPY expression, shown in f.

TABLE 1 Effect of intracerebroventricular cumulative MCH on chow consumed (grams)

Experiment		1 Hour	P	1.5 Hours	P	2 Hours	P	4 Hours	P	6 Hours	P
1	Vehicle	1.4 ± 0.4	0.043			2.3 ± 0.4	0.001	4.5 ± 0.6	0.021	6.3 ± 0.7	n.s.
	MCH	3.2 ± 0.7				5.2 ± 0.5		6.7 ± 0.6		7.7 ± 0.9	
2	Vehicle	2.2 ± 0.4	n.s.			5.4 ± 0.4	0.054	7.7 ± 0.7	0.0304	10.6 ± 0.3	0.0083
	MCH	5.4 ± 1.6				9.3 ± 1.9		12.3 ± 1.7		16.0 ± 1.3	
3	Vehicle	2.3 ± 0.6	n.s.			3.3 ± 0.8	n.s.	4.4 ± 0.8	0.0077	5.5 ± 0.9	0.0121
	MCH	2.6 ± 0.4				4.7 ± 0.3		6.8 ± 0.3		8.6 ± 0.8	
4	Vehicle			2.2 ± 0.6	n.s.			3.0 ± 0.9	0.0108		
	MCH			4.0 ± 1.2				6.7 ± 0.9			
5	Vehicle			4.2 ± 0.3	0.068			5.4 ± 0.5	0.0202		
	MCH			8.3 ± 1.8				10.7 ± 1.6			

Experiment 1: 1.5 µg MCH in artificial cerebrospinal fluid (CSF) was injected into eight Long-Evans rats through an indwelling intraventricular catheter. The control group consisted of eight rats injected with artificial CSF only. All rats were housed individually under reversed light-dark cycle, and injected at 9:00 a.m., which was the beginning of the dark period. Food intake (ground Purina Laboratory Chow) was measured at 1, 2, 4 and 6 h after injections. Administration of MCH led to an increase in eating behaviour; at 1 and 2 h after injections food intake of MCH-treated animals was more than twice that of control animals ($P = 0.043$) and $P = 0.001$, respectively). At 4 h after injections food intake of MCH-treated animals was 46% increased over control animals ($P = 0.021$). After 6 h, cumulative food intake of MCH-treated rats was greater than controls but the difference was not significant ($P = 0.22$). Each data point represents the mean grams of chow $\pm$ the standard error of eight animals. Experiments 2 and 3 were done under similar conditions. Experiments 4 and 5 were at the end of the dark cycle in rats habituated to a mash diet (27% sugar, 27% condensed milk, 45% ground rat chow). In addition, in experiment 5, a higher dose (30 µg) of MCH was administered. Although an increase in feeding behaviour was seen in all time points in all experiments there was some experimental variation, particularly at the earlier time points. However at 4 h MCH caused a statistically significant increase in caloric consumption in all experiments. Stainless steel outer guide cannulas (23 gauge; Plastic Products, Roanoke, VA) were implanted into the right lateral ventricle in male Long-Evans rats using a Kopf stereotaxic instrument. Surgery was performed in rats anaesthetized with 100 mg kg⁻¹ ketamine hydrochloride and 6 mg kg⁻¹ Xylazine (i.p.) in a sterile field using autoclaved instruments. The head was shaved and washed with alcohol and Betadine. The animal was placed into the stereotaxic instrument and its head held with two earbars and an incisor bar. A midline incision was made along the length of the skull and the skin and connective tissue were retracted with serrated clips. With the skull level between bregma and lambda, the cannula was implanted into the right lateral ventricle (1.0 mm posterior to bregma, 1.6 mm lateral to the midline, and 4.0 mm ventral to the surface of the dura), and secured with Ketac cranial cement. The incision was sutured with sterile silk, and a 30-gauge stainless steel stylette placed in the cannula to keep it patent. The animal had 20 mg kg⁻¹ Keflin (im) at the time of surgery and one daily injection of 20 mg kg⁻¹ Keflin for 3 days immediately following surgery. Animals were allowed to recover for at least 2 weeks before intraventricular injections. Animals were handled on a daily basis for cannula maintenance. Injections of 5 µg synthetic rat MCH (Bachem, Switzerland) were made in a total volume of 5 µl. For the first 6 h after injection, animals were observed continuously and no changes in behaviour were noted. At the conclusion of the experiment the animals were killed with sodium pentobarbital (100 mg kg⁻¹, i.p.) and the brain was obtained for histological verification of catheter placement.

a 6-h period. This increase is similar to that seen after galanin administration and somewhat less than seen after NPY treatment⁸. Because ICV administration does not entirely mimic the normal distribution of neuropeptides, and because MCH is known to have diffuse direct projections into both cortex and brainstem, the relative magnitude of the feeding effects should not be over-interpreted.

MCH, a cyclic, 19-amino-acid neuropeptide, was initially identified in the intermediate lobe of teleost fish pituitary from which it is released into the circulation and causes aggregation of melanophores^{9,10}. In mammals, cell bodies containing immunoreactive MCH are found exclusively in the ventral aspect of the zona incerta and the lateral hypothalamus¹¹⁻¹³ and expression of the MCH mRNA is confined to these areas of the brain^{7,11,14}. The MCH gene encodes MCH, as well as two other peptides¹⁵. One, a 13-amino-acid peptide designated neuropeptide EI, shares epitopes with CRF^{6,16} and is processed and released by hypothalamic cells in culture¹⁷. The other encoded sequence is for a 19-amino-acid peptide, neuropeptide-GE, which has sequence homology to human growth-hormone-releasing factor¹⁸.

The physiological function of MCH in mammals is unknown, but a potential role in the integrative processes that accompany complex behaviours¹⁹ has been postulated on the basis of the location of MCH neuronal bodies and their projections throughout the mammalian brain.

The potential relationship between MCH and melanin-stimulating hormone (MSH) is noteworthy. First, although there is no amino-acid sequence similarity, MCH and MSH may share a common epitope in that anti-MSH antibodies are reported to crossreact with neurons containing MCH²⁰. Second, MCH antagonizes the action of α -MSH in the skin of teleost fish²¹, whereas high concentrations of MCH in reptiles and amphibians have an MSH-like action²², presumably as a result of interaction with one

or more of the melanocortin receptors²³. These intriguing connections between MCH and MSH are made relevant to the biology of obesity by recent observations regarding the mechanism of obesity in the A^{vy} yellow mouse^{24,25}. Obesity in these mice is due to ectopic expression of agouti protein, which binds to melanocortin receptors and antagonizes the action of MSH²⁶. Thus, one mechanism by which the agouti protein may cause obesity would be to mimic the action of the orexigenic peptide MCH in the hypothalamus.

Because MCH expression in the hypothalamus is increased by fasting and in the *ob/ob* mouse, and because ICV MCH administration stimulates food intake, this neuropeptide must be viewed as a potentially important effector of nutritional homeostasis. Future experiments will determine whether MCH has additional actions relevant to nutritional homeostasis, such as effects on the regulation of energy expenditure and the secretion of metabolic hormones such as insulin. It will also be necessary to define the relationship between MCH and other central regulators of energy homeostasis, and potential interactions between MCH and peripheral regulators of energy homeostasis such as the fat-derived hormone leptin^{27,28}. □

Received 23 October 1995; accepted 26 January 1996.

- Stanley, B. G., Anderson, K. C., Grayson, M. H. & Leibowitz, S. F. *Physiol. Behav.* **46**, 173-177 (1989).
- Kyrkouli, S. E., Stanley, B. G. & Leibowitz, S. F. *Eur. J. Pharmacol.* **122**, 159-160 (1986).
- Arase, K., York, D. A., Shimizu, H., Shargill, N. & Bray, G. A. *Am. J. Physiol.* **255**, E255-E259 (1988).
- Turton, M. D. et al. *Nature* **379**, 69-72 (1996).
- Liang, P. & Pardee, A. B. *Science* **257**, 967-971 (1992).
- Nahon, J. L., Presse, F., Breton, C., Hervieu, G. & Schorpp, M. *Ann. N.Y. Acad. Sci.* **680**, 111-129 (1993).
- Breton, C., Presse, F., Hervieu, G. & Nahon, J. L. *Molec. cell. Neurosci.* **4**, 271-284 (1993).
- Tempel, D. L. & Leibowitz, S. F. *Brain Res. Bull.* **25**, 821-825 (1990).
- Kawauchi, H., Kawazoe, M., Tsubplawa, M., Kishida, M. & Baker, B. I. *Nature* **305**, 423-432 (1983).

10. Kawauchi, H., Naito, N. & Ono, M. *Ann. N.Y. Acad. Sci.* **680**, 64–77 (1993).
11. Skofitsch, G., Jacobowitz, D. M. & Zamir, N. *Brain Res. Bull.* **15**, 635–639 (1985).
12. Zamir, N., Skofitsch, G., Bannon, M. J. & Jacobowitz, D. M. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1528–1531 (1986).
13. Naito, N., Kawazoe, I., Nakai, Y. & Kawauchi, H. *Cell Tissue Res.* **253**, 291–295 (1988).
14. Thompson, R. C. & Watson, S. J. *DNA Cell Biol.* **9**, 637–645 (1990).
15. Bittencourt, J. C. et al. *J. comp. Neurol.* **319**, 218–245 (1992).
16. Nahon, J. L., Presse, F., Bittencourt, J. C., Sawchenko, P. E. & Vale, W. *Endocrinology* **125**, 2056–2065 (1989).
17. Parkes, D. G. & Vale, W. *Endocrinology* **131**, 1826–1831 (1992).
18. Risold, P. Y., Fellmann, D., Rivier, J., Vale, W. & Bugnon, C. *Neurosci. Lett.* **136**, 145–149 (1992).
19. Oomura, Y. *New Physiol. Sci.* **2**, 199–203 (1987).
20. Cechetto, D. F. & Saper, C. B. *J. comp. Neurol.* **272**, 579–604 (1988).
21. Rance, T. & Baker, B. I. *Gen. comp. Endocrin.* **37**, 64–73 (1979).
22. Matsunaga, T. O., Hruby, V. I., Lebl, M., De Lauro Castrucci, A. M. & Hadley, M. E. *Peptides* **10**, 773–778 (1989).
23. Cone, R. D. et al. *Ann. N.Y. Acad. Sci.* **80**, 342–363 (1993).
24. Bultman, S. J., Michaud, E. J. & Woychik, R. P. *Cell* **75**, 1195–1204 (1992).
25. Klebrig, M. L., Wilkinson, J. E., Geisler, J. G. & Woychik, R. P. *Proc. natn. Acad. Sci. U.S.A.* **92**, 4728–4732 (1995).
26. Lu, D. et al. *Nature* **371**, 799–803 (1994).
27. Zhang, Y. et al. *Nature* **372**, 425–432 (1994).
28. Pellemounter, M. A. et al. *Science* **269**, 540–543 (1995).
29. Jakubowski, M. & Roberts, R. L. *J. Neuroendocrin.* Appendix 2 (1991).
30. Sanacora, G., Kershaw, M., Finkelstein, J. A. & White, J. D. *Endocrinology* **127**, 730–736 (1990).

ACKNOWLEDGEMENTS. D.Q. and D.S.L. contributed equally to this work. We thank C. R. Kahn and D. Vicent for discussions; J. S. Flier for support and advice; J. Marron for secretarial assistance; and L. Aiello for advice on PCR display. This work was supported in part by a grant from the NIH and by the Lucille P. Markey Charitable trust which also partly supported E.M.F. D.S.L. was supported in part by a fellowship from the Lawson Wilkins Pediatric Endocrine Society and by an NIH grant to J. Flier. This work was also supported in part by a grant from the National Institute on Drug Abuse to R.K. Most sequencing reactions were performed by the Joslin Diabetes Center DNA core laboratory, which is supported in part by the Lucille P. Markey Charitable Trust. Animal costs were supported in part by the Animal Core of DERC.

Learning to commit or avoid the base-rate error

Adam S. Goodie & Edmund Fantino

Department of Psychology, University of California, San Diego, La Jolla, California 92093, USA

WHEN predicting an event, people neglect overall frequencies (base rates) of various possibilities¹. We have previously shown² that this base-rate occurs not only with word problems, but also in a procedure with repeated trials: a sample cue followed by two choice options, one of which the subject must choose, with feedback regarding the correctness of the choice. Perhaps this base-rate error depends on people's histories of matching physically similar items. In support of this suggestion, we begin by showing that the base-rate error is eliminated with physically unrelated items. We then show that when the relation between items is again arbitrary, but is a relation that subjects already know, the error reappears. Finally we show that teaching subjects new arbitrary relations reintroduces the error in later testing. These experiments demonstrate a fundamental base-rate error dependent on learned relationships: without interference from pre-existing associations between cues and options, predictions may be made more optimally.

When assessing the probability of a future event, people often ignore background information in favour of case-specific evidence, an effect called the base-rate error^{1,3}. Consider this problem⁴:

A cab was involved in a hit and run accident at night. Two cab companies, the Green and the Blue, operate in the city. You are given the following data:

- (a) 85% of the cabs in the city are Green and 15% are Blue.
- (b) a witness identified the cab as Blue... The witness correctly identify[s] each one of the two colors 80% of the time and fail[s] 20% of the time. What is the probability that the cab involved in the accident was Blue rather than Green?

It is more common for the witness to see a green taxicab and mistakenly call it blue ($0.85 \times 0.20 = 17\%$ of all cases) than for the witness to see a blue taxi and label it correctly

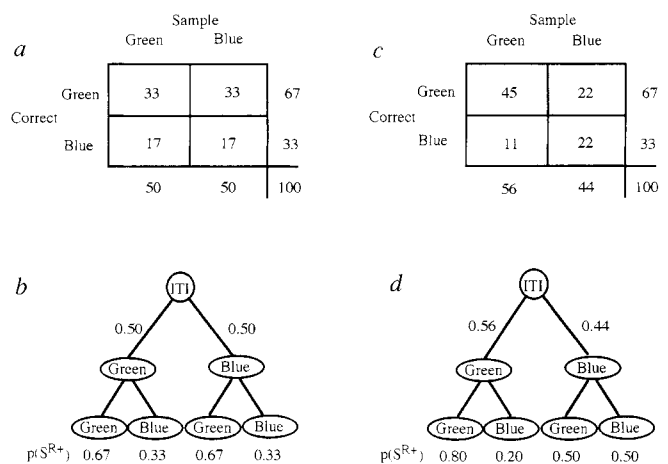


FIG. 1 Procedure used in prior experiments². **a**, An incidence table derived from a taxi problem with a base rate of 67% green and 33% blue, a cue accuracy of 50%, and a sample of 100 total instances. Green is the correct response in 67% of cases, and 50% of those (33) are correctly marked by a green cue, while the other 50% (33) are marked by a blue cue. Of the 33% of all instances where blue is the correct response, 50% (17) are correctly marked by a blue cue, while the other 50% (17) are marked by a green cue. **b**, The conversion of this table to a cue-matching experimental design. Because the green cue (G) appears 50 (33 + 17) times out of every 100, the green cue is presented 50% of the time, and the blue cue (B) appears the remaining 50%, after an inter-trial interval (ITI). When green has appeared, green is correct 33/50 = 67% of the time, and blue is correct the other 33% of the time; these are indicated by their probabilities of reinforcement ($p(S^R)$). Hence, after a green cue, choosing green is reinforced 67% of the time, and choosing blue is reinforced 33% of the time. By identical reasoning, when a blue cue appeared, choosing green is reinforced 67% of the time, and choosing blue is reinforced 33% of the time. Note that the 50%-accurate cue is irrelevant to the contingencies of reinforcement. **c**, **d**, The same functions as **a** and **b**, respectively, for the condition where the base rate was still 67%, but where the cue was 67% accurate. Procedural details are as previously described².

($0.15 \times 0.80 = 12\%$ of all cases). If the witness reports seeing a blue taxi, the probability that the cab actually was blue is $0.12 / (0.17 + 0.12) = 41\%$. This is one application of Bayes's theorem, which is often held as an optimal model of behaviour. Subjects responding to the 'taxicab problem', however, ignore base rates (how many green and blue taxis operate in the city) and rely on the reliability of what the witness says.

We have found the base-rate error in a different sort of setting². On each of many trials, subjects predicted whether the 'correct' answer to a problem would be green or blue. In 67% of cases, the correct answer was green, with blue being correct the other 33% of the time. On each trial a cue was presented that anticipated the correct answer with either 50% or 67% accuracy. We measured how often subjects matched the cue, guessing green if a green cue had appeared, or blue if a blue one appeared (see Fig. 1).

Two particular data patterns would represent a base-rate error. One is roughly equal matching to the blue and green cues, indicating that subjects treat both cues alike despite differences in their base rates. This failure to distinguish between cue types is a characteristic feature of the base-rate error. The other is pre-dominant matching of the blue cue, which under some conditions is distinctly suboptimal. The results of previous experiments appear in Fig. 2, and show both tell-tale patterns: subjects matched blue more than half the time, and the average difference between matching green and blue was only 7%.

We propose that the base-rate error is a learned phenomenon, and present three supporting experiments. The first two differ from our previous design only in the physical nature of the cues. Experiment 1 used cues that bore no natural or previously learned relation to the choice options. If the base-rate error results from

subjects' histories of matching items that are similar to one another, then this should reduce the base-rate error. The green and blue cues were replaced respectively by vertical and horizontal lines, and the base-rate error disappeared entirely (see Fig. 3a). The 50% condition, in which subjects had previously lost points by preferring blue following a blue cue, now showed a marked preference for green following the horizontal cue. Even in the 67% group, where strategy following a 'blue' cue was irrelevant, the proportion of times it was matched fell considerably, removing the characteristic flatness of slope. The difference between matching 'green' and matching 'blue', averaged across groups, increased from 7% to 28%.

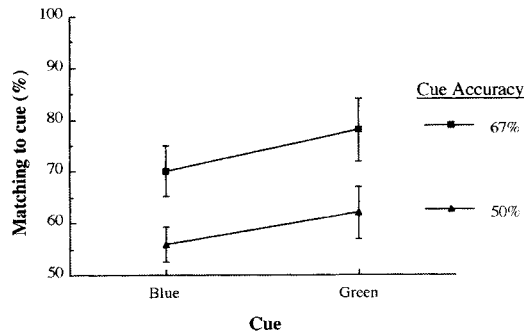
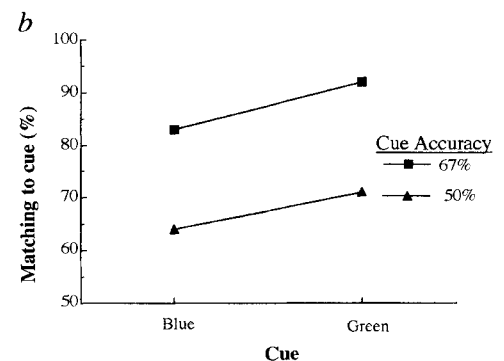
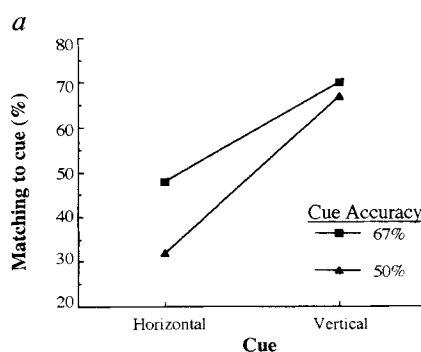


FIG. 2 Data from prior experiment². The base-rate error is represented by two graphical features: the high matching of the blue cue in both groups,

FIG. 3 Data from Experiments 1 and 2. *a*, The base-rate error disappeared in Experiment 1: matching of the 'blue' cue was much lower than it has been in previous work², and the slopes of lines increased markedly, being significantly greater than zero ($F(1, 15) = 28.09$, $P < 0.05$). (The 'blue' cue was in fact not blue but a white horizontal line. We continue to call this matching 'blue' to clarify the relationship of this experiment to previous ones² and to Experiments 2 and 3: horizontal lines replace blue cues here, and vertical lines replace green cues. In all experiments the choice options were blue and green patches, presented side-by-side on a computer monitor.) The increase in slope is partly obscured because the ordinate axis is altered to accommodate the increased range of values. *b*, The base-rate error reappeared in Experiment



2. Matching of the 'blue' cue returned to levels at or above those observed in previous studies², and slopes returned to near-zero, although they were once again significant ($F(1, 10) = 6.28$, $P < 0.05$).

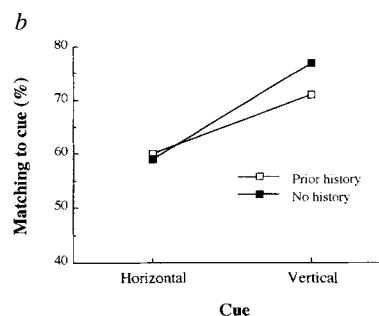
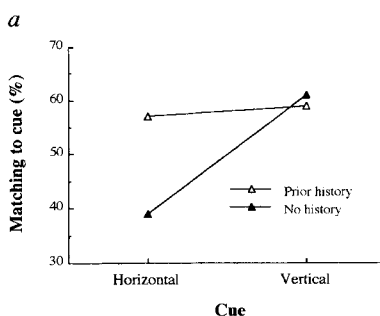


FIG. 4 Data from Experiment 3. Subjects in the 'Prior history' condition were first exposed to 200 trials in which green was always correct following a vertical cue, and blue was always correct following a horizontal cue. For these subjects, a horizontal line was thus established as equivalent to a blue cue, and so choosing blue following a horizontal line continues to be called matching 'blue'. Those in the 'No history' condition saw only the 200 trials of the test phase, which was identical to Experiment 1. Subjects were divided into two levels of cue accuracy, 50% and 67%, depicted in *a* and *b*, respectively. The 'Prior history' groups show substantially flatter slopes, resulting in cross-over effects at both levels of cue accuracy, which are statistically significant ($t(22) = 2.04$, $P < 0.05$).

correct, and between a vertical cue and green being correct. This induced our subjects to a greater flatness of slope than those without prior training, a distinct trend towards the base-rate error (see Fig. 4). The average difference between matching 'green' and matching 'blue' was 20% for subjects for no prior training, but only 6% for those with prior training.

Base-rate neglect has previously been obtained⁵⁻⁷ in experimental settings. However, this has generally involved cues and choices with some prior relationship, such as symptom and disease, but not in as dependable a relationship as identity. By using elements with either no prior relation or a lock step relation of identity, we have isolated the effects of what is previously known (Experiments 2 and 3) from that which is learned during an experimental session (Experiments 1 and 3). These experiments demonstrate a base-rate error that depends on learned probabilistic relationships: pre-existing associations between cue and outcome can prevent learning from experience. □

Received 28 August 1995; accepted 8 January 1996.

1. Bar-Hillel, M. *Acta psychol.* **44**, 211–233 (1980).
2. Goodie, A. S. & Fantino, E. *Psychol. Sci.* **6**, 101–106 (1995).
3. Kahneman, D. & Tversky, A. *Psychol. Rev.* **80**, 237–251 (1973).
4. Tversky, A. & Kahneman, D. in *Judgment under Uncertainty: Heuristics and Biases* (eds Kahneman, D., Slovic, P. & Tversky, A.), 153–160 (Cambridge Univ. Press, 1982).
5. Gluck, M. A. & Bower, G. H. *J. exp. Psychol., gen.* **117**, 227–247 (1988).
6. Medin, D. L. & Edelson, S. M. *J. exp. Psychol., gen.* **117**, 63–85 (1988).
7. Nosofsky, R. M., Kruschke, J. K. & McKinley, S. C. *J. exp. Psychol., Learn. Memory Cogn.* **18**, 211–233 (1992).
8. Chapman, L. J. & Chapman, J. P. *J. abnorm. Psychol.* **72**, 193–204 (1967).
9. Chapman, L. J. & Chapman, J. P. *J. abnorm. Psychol.* **74**, 271–280 (1969).
10. Myers, J. L. in *Handbook of Learning and Cognitive Processes* Vol. 3 (ed. Estes, W. K.) 171–205 (Erlbaum, Hillsdale, NJ, 1976).
11. Herrnstein, R. J. & Loveland, D. H. *J. exp. Analysis Behav.* **24**, 107–116 (1975).
12. Hartl, J. A. & Fantino, E. *J. exp. Analysis Behav.* (in the press).

ACKNOWLEDGEMENTS. We thank J. Otsuka for assistance in collecting data. This work was supported by a grant from the NSF.

Orientation selectivity of thalamic input to simple cells of cat visual cortex

David Ferster, Sooyoung Chung & Heidi Wheat

Department of Neurobiology and Physiology, Northwestern University, 2153 North Campus Drive, Evanston, Illinois 60208, USA

MORE than 30 years after Hubel and Wiesel¹ first described orientation selectivity in the mammalian visual cortex, the mechanism that gives rise to this property is still controversial. Hubel and Wiesel¹ proposed a simple model for the origin of orientation tuning, in which the circularly symmetrical receptive fields of neurons in the lateral geniculate nucleus that excite a cortical simple cell are arranged in rows. Since this model was proposed, several experiments²⁻⁶ and neuronal simulations^{7,8} have suggested that the connectivity between the lateral geniculate nucleus and the cortex is not well organized in an orientation-specific fashion, and that orientation tuning arises instead from extensive interactions within the cortex. To test these models we have recorded visually evoked synaptic potentials in simple cells while cooling the cortex⁹, which largely inactivates the cortical network, but leaves geniculate synaptic input functional. We report that the orientation tuning of these potentials is almost unaffected by cooling the cortex, in agreement with Hubel and Wiesel's original proposal¹.

Whole-cell patch recordings were obtained from 10 simple cells in layer 4 of the cat primary visual cortex. Orientation selectivity was measured from the responses to drifting sinusoidal luminance

gratings, which evoked approximately sinusoidal modulations of the membrane potential with peak-to-peak amplitudes of up to 25 mV. Once the normal tuning properties of each cell were characterized, the cortex was cooled.

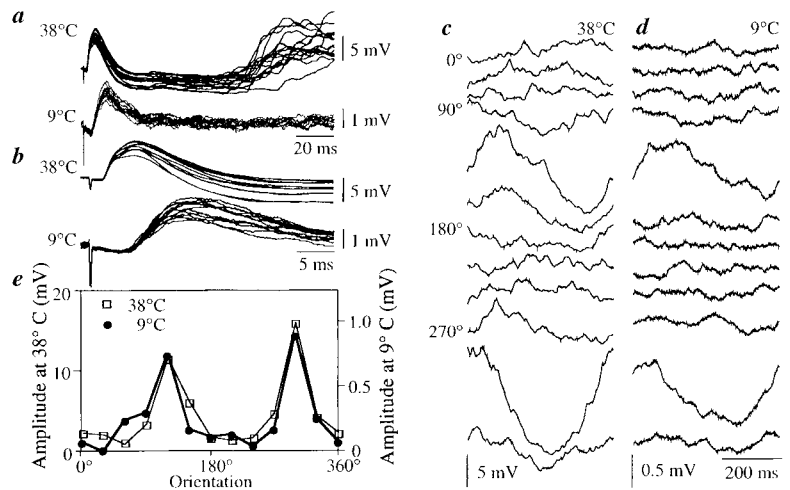
During cooling, seven different aspects of the responses of layer 4 cells to electrical and visual stimulation changed. The temperature at which these changes occurred (4–14 °C) varied from cell to cell, presumably depending on the depth of recording. 1) Visually and electrically evoked action potentials were gradually reduced in number until they disappeared entirely. Other neurons within layer 4 and in layers closer to the cooling plate were also likely to have been silenced. 2) Disynaptic inhibitory postsynaptic potentials (i.p.s.ps) evoked by electrical stimulation of the lateral geniculate nucleus (LGN) disappeared (Fig. 1a). Before cooling, the i.p.s.p. peak amplitude averaged 6 mV; during cooling the average amplitude was reduced by 95% to 0.28 mV. In six cases, cooling completely suppressed the i.p.s.p. and the membrane potential never dipped below rest. It appears, then, that during cooling, most inhibitory interneurons no longer responded to electrical stimulation. 3) Most excitatory interneurons were also silenced. During cooling, all visually and electrically evoked synaptic potentials, including excitatory postsynaptic potentials (e.p.s.ps), disappeared in four of the five complex cells recorded, all of which lacked monosynaptic excitation from the LGN. 4) The electrically evoked monosynaptic e.p.s.p. gradually increased in latency by a factor of 2–3 (Fig. 1b), presumably as a result of slowed synaptic release in geniculate afferent terminals. The latency of the augmenting response, a monosynaptic potential mediated by antidromic activation of layer-6 corticogeniculate neurons, increased along with the latency of the geniculocortical input (Fig. 2). 5) The electrically evoked monosynaptic e.p.s.ps were reduced in amplitude (Fig. 1a); the average reduction in 10 cells was to 40% of normal. 6) Visually evoked responses were also reduced in amplitude (Fig. 1c, d), but this reduction was proportionately greater than the reduction in the electrically evoked e.p.s.p. In addition, the noise of the traces was sharply reduced so that the signal-to-noise ratio was hardly changed (compare Fig. 1c, d, and note change in vertical scale), suggesting that a large component of the noise was synaptic in origin. 7) The resting input resistance of the recorded cell approximately doubled, presumably because of the reduction in spontaneous synaptic activity. Taken together, these changes indicate that cortical activity was significantly depressed by cooling, but that the geniculate input remained functional.

Although cortical activity in layer 4 and above was strongly suppressed by cooling, cells in layer 6, being farther from the cooling plate, may have been warm enough to remain active. These cells provide strong excitatory input to layer 4 (refs 10, 11), and their synapses, although weakened, still functioned, as indicated by the response to electrical stimulation of the LGN (Fig. 2b). To determine whether the layer-6 cells remained visually responsive during cooling, we recorded extracellularly from three sites in layer 6. At the cooled temperatures used in the intracellular experiments, visually evoked multi-unit activity fell by an average of 85%. Considering the strong temporal facilitation observed in the synapses between layer-6 and layer-4 cells¹¹, a six-fold decrease in activity in layer 6 will result in a much larger reduction in synaptic input to layer-4 cells. This reduction, when compounded with the direct effect of cooling on the layer-6 cell's synapses in layer 4 (Fig. 2a, b, downward arrows), should result in a 25-fold or greater reduction in the amplitude of the input from layer 6. We conclude from these observations that cooling severely disrupted activity in the entire local cortical circuit underneath the cooling plate.

With activity in the cortical circuit strongly suppressed, the orientation tuning of the remaining monosynaptic geniculate input to simple cells could be studied. If the model proposed by Hubel and Wiesel¹ is correct, synaptic input should be as well tuned in the cold as at normal temperatures. On the contrary, the cortically based models predict that the response of the geniculate

FIG. 1 The effects of cooling on the visually and electrically evoked responses of a layer-4 simple cell. *a*, response of a simple cell to electrical stimulation of the LGN (1 mA, 200 μ s) recorded at normal temperature and during cooling. *b*, Similar traces shown at a different time scale. The short (1.8 ms) and stable latency of the e.p.s.p. indicates that this cell received monosynaptic excitation from the LGN. The monosynaptic e.p.s.p. was followed by a long-lasting (150 ms) i.p.s.p. of disynaptic origin. When the cooling plate temperature was lowered to 9 °C, the latency of the e.p.s.p. increased to 5 ms, its rise time was dramatically slowed, and its amplitude was reduced; the i.p.s.p. disappeared. *c*, The cell's response to drifting gratings of 12 different orientations. Each trace is the averaged response to 20 grating cycles. *d*, Responses to the same visual stimuli as in *c*, but with the cortex cooled. The responses are similar in shape to those in *c*, but more than 17-fold smaller in amplitude (note different vertical scales in the figure). *e*, Orientation tuning curves constructed from the records of *c* and *d*. Each point indicates the peak-to-peak amplitude of the first harmonic (2 Hz) component of the corresponding trace. Once again, note the different vertical scales for the two plots.

METHODS. Whole-cell patch recordings in current-clamp mode were obtained from barbiturate-anaesthetized, paralysed adult cats¹⁸. The cortex was cooled by an L-shaped copper plate^{9,30}. The horizontal portion measured 4 × 4 mm and rested on the cortical surface near the area 17/18 border between 2 and 6 mm posterior of Horsley-Clarke 0. A hole (0.9 mm in diameter) in the cooling plate allowed the insertion of electrodes into the cortex. A thermocouple embedded in the top surface of the plate indicated the plate temperature. The vertical portion of the plate was pressed against a Peltier device. In some experiments a second L-shaped copper plate rested on the visual cortex of the opposite hemisphere. Warm agar (3% in physiological saline) sealed the cooling plate to the surrounding bone and



sealed the electrode in place. Drifting sinusoidal gratings of optimal spatial frequency, 2 Hz temporal frequency, and 64% contrast were presented monocularly on an oscilloscope screen (mean luminance of 15 cd m⁻²) using an image generator (Innisfree, Cambridge, MA). Electrical stimuli were delivered through a sharpened tungsten electrode to the optical radiations just overlying the LGN. Layer-4 simple cells were identified by the presence of two or three subregions in their receptive fields¹, and by their characteristic response to electrical stimulation of the LGN¹¹. Layer-6 corticogeniculate cells were identified in extracellular recordings by antidromic activation from the LGN and by subsequent histology.

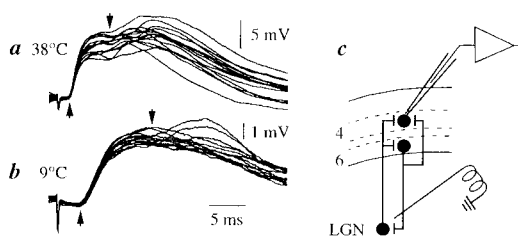
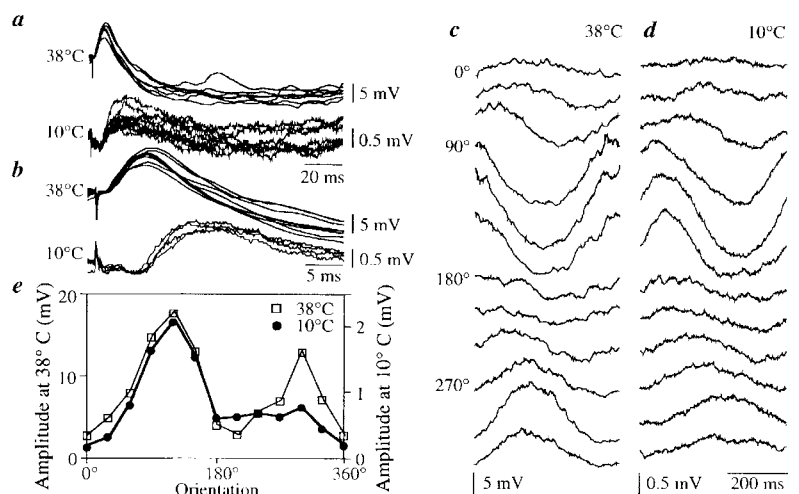


FIG. 2 The effect of cooling on the latency of electrically evoked monosynaptic potentials. *a*, Response to electrical stimulation of the LGN at 10 Hz. At this high frequency, a long-latency e.p.s.p. component (downward arrow) rises out of the monosynaptic response to geniculate relay cell activation (upward arrow). It reflects a monosynaptic potential mediated by the antidromic activation of layer-6 corticogeniculate neurons and their collaterals in layer 4. Its long latency is a result of the slowly conducting corticogeniculate axons¹¹. *b*, When the cortex was cooled, the latencies of the two components increased by a similar factor. *c*, A diagram of the connections between the cortex and the LGN that mediate the two potentials seen in *a* and *b*.

FIG. 3 The effects on the visually and electrically evoked responses of a second simple cell. The layout is identical to that of Fig. 1. Slight shifts in the temporal phase of some of the visually evoked responses were apparent during cooling (compare *c* and *d*). These shifts were not present in the cell shown in Fig. 1. They were probably caused by small (<0.5°) shifts in the position of the eyes; in other cells, similar shifts over time were observed independently of cooling. In this cell, cooling increased direction selectivity.



input alone should be poorly tuned for orientation compared with the total synaptic input. Our data from layer 4 conform more to the prediction of Hubel and Wiesel's model¹: the amplitude of visually evoked responses declined during cooling (Fig. 1e, note different vertical scales), but the orientation tuning remained virtually identical. In addition, the overall time course of individual responses changed little (Fig. 1c, d). When tested, rewarming restored the responses nearly to their original amplitudes. The effects of cooling on a second neuron are shown in Fig. 3, and a summary of the effects of cooling on the ten neurons in our sample is shown in Fig. 4. The consistent broadening of orientation tuning predicted by the cortically based models was not apparent.

The data in Fig. 4 may be used to estimate the contribution of thalamic inputs to the visually evoked responses recorded at normal temperatures. We first assume, from the disappearance of polysynaptic postsynaptic potentials, that all but the thalamic component of the visually evoked response are virtually eliminated by cooling. We then assume that cooling reduces the amplitude of the thalamic component of the visually evoked responses and the amplitude of the electrically evoked monosynaptic e.p.s.p. by the same factor. If our assumptions are correct then the amplitude of the thalamic component of the visually evoked response at normal temperatures will equal the amplitude of the cooled visually evoked response divided by this factor. If R is the amplitude of a given response, then

$$R_{\text{visual, warm}}(\text{geniculate component}) = R_{\text{visual, cold}} \times \frac{R_{\text{electrical, warm}}}{R_{\text{electrical, cold}}}$$

This calculation applied to Fig. 4 indicates that the LGN contributes between 21% and 63% of the visually evoked responses. The mean contribution (37%) matches well with direct autoradiographic measurements of the proportion of excitatory terminals in layer 4 that originate in the LGN (28%) (ref. 12).

Although cooling the cortex surrounding the recording site strongly inactivated the local circuit, three remote sources of orientation-selective input in addition to the LGN might have remained active: area 18 (ref. 10); uncooled parts of area 17 projecting by means of long-range horizontal connections¹³; and areas 17 and 18 of the opposite hemisphere^{14,15}. These projections largely target neurons outside layer 4. Furthermore, the suppression of nearly all measurable visually evoked activity in neurons without direct input from the LGN indicates that these sources do not provide significant input during cooling. Nevertheless, three procedures were conducted to control for the possible effects of these inputs: to limit input from area 18, all neurons were recorded near the border between areas 17 and 18, so that the cooling plate spanned both areas; to limit input from uncooled parts of area 17, for four cells (Fig. 4a–d) the visual stimulus, normally 8° in diameter, was limited in area to no more than twice the diameter of the receptive field; and to limit input from the opposite hemisphere, for three cells (Fig. 4b–d) the cooling plate was enlarged to cover corresponding regions of areas 17 and 18 in both hemispheres. None of these manipulations significantly affected the results.

Early cortically based models of orientation selectivity relied on lateral inhibition among neurons with different preferred

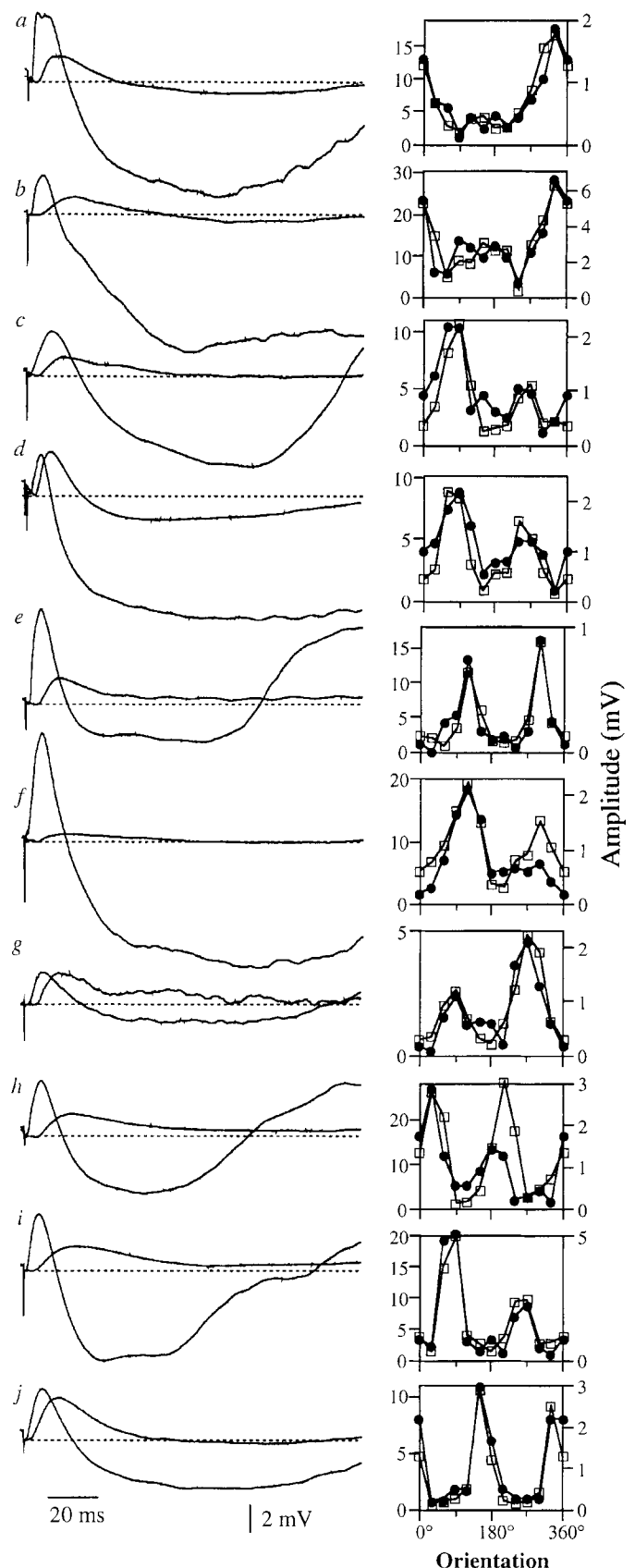


FIG. 4 Left, averaged responses to electrical stimulation of the LGN for each of the ten cells in our sample. In each case, the larger of the two responses was recorded at normal temperature and the smaller during cooling. V_{rest} is indicated by the dotted lines. Right, orientation tuning curves for each cell at normal and cool temperatures. Responses recorded at 38 °C (open squares) are plotted at the scale to the left of each tuning curve. Responses recorded at cool temperatures (filled circles) are plotted at the scale to the right of each tuning curve. In a–d, the visual stimulus was restricted to twice the size of the receptive field. In b–d, the cortex was cooled bilaterally. Cool temperatures for the 10 cells were: a, 5 °C; b, 14 °C; c, 9 °C; d, 4 °C; e, 9 °C; f, 10 °C; g, 6 °C; h, 14 °C; i, 8 °C; j, 10 °C.

orientations². These models were supported by the observation that removal of inhibition by extracellularly applied GABA_A antagonists broadened or even abolished orientation selectivity⁴. Intracellular records, however, failed to reveal the significant inhibition that was predicted by lateral-inhibition models to be evoked by stimuli at the non-preferred orientation^{16–18}. In addition, intracellular blockade of inhibition within a single neuron did not change orientation selectivity¹⁹. As a result, it has been suggested that orientation selectivity might arise not only from cortical inhibitory interactions, but from selective amplification of geniculate inputs by excitatory feedback within the cortical column^{7,8,20}. The cooling experiment shows, however, that severely disrupting interactions within the cortical column, including excitatory feedback, leaves orientation tuning intact. In layer 4, therefore, the synaptic input from the LGN that remains during cooling is sufficient to determine orientation selectivity.

Similar questions arise concerning the origin of direction selectivity. Saul and Humphrey²¹ suggested that direction selectivity is generated from temporally and spatially offset input from lagged and non-lagged geniculate neurons. On the other hand, the cortical amplification mechanisms that were applied to the problem of orientation selectivity can also be adapted to direction selectivity^{7,20,22,23}. Our results are consistent with Saul and Humphrey's model²¹. In no case did cooling significantly decrease direction selectivity (Fig. 4). In two cases, cooling increased direction selectivity (Fig. 4f, h), as if the cortical inputs were supplying excitation in the non-preferred direction. The basis for direction selectivity, like orientation selectivity, may therefore be established by geniculate input.

As predicted by Hubel and Wiesel¹, on and off subregions do seem to originate from monosynaptic input from on- and off-centre geniculate cells^{24,25}. It has been argued²⁶ from precise extracellular measurements of the shapes of simple cell on and off subregions that the elongation of these subregions is often sufficient to account for orientation selectivity. An anatomical substrate for the formation of elongated subfields by geniculate has been suggested²⁷, as the receptive fields of geniculate afferents innervating a small region of cortex are arranged along lines parallel to the preferred orientation of neurons within the column²⁷. It now remains to be determined how this complex organization of geniculate inputs develops during the prenatal period in the absence of visual input^{28,29}. □

Received 27 October 1995; accepted 19 January 1996.

- Hubel, D. H. & Wiesel, T. N. *J. Physiol., Lond.* **160**, 106–154 (1962).
- Blakemore, C. & Tobin, E. A. *Exp. Brain Res.* **15**, 439–440 (1972).
- Creutzfeldt, O. D., Kuhn, U. & Benevento, L. A. *Exp. Brain Res.* **21**, 251–274 (1974).
- Sillito, A. M. *J. Physiol., Lond.* **250**, 305–329 (1975).
- Crook, J. M., Eysel, U. T. & Machemer, H. F. *Neuroscience* **40**, 1–12 (1991).
- Kisvarday, Z. F., Kim, D. S., Eysel, U. T. & Bonhoeffer, T. *Eur. J. Neurosci.* **6**, 1619–1632 (1994).
- Douglas, R. J. & Martin, K. A. C. *J. Physiol., Lond.* **440**, 735–769 (1991).
- Somers, D. C., Nelson, S. B. & Sur, M. *J. Neurosci.* **15**, 5448–5465 (1995).
- Kail, R. E. & Chase, R. J. *Neurophysiol.* **33**, 459–474 (1970).
- Gilbert, C. D. & Kelly, J. P. *J. comp. Neurol.* **163**, 81–105 (1975).
- Ferster, D. & Lindström, S. *J. Physiol., Lond.* **367**, 233–252 (1985).
- LeVay, S. & Gilbert, C. D. *Brain Res.* **113**, 1–19 (1976).
- Gilbert, C. D. & Wiesel, T. N. *Nature* **280**, 120–125 (1979).
- Hubel, D. H. & Wiesel, T. N. *J. Neurophysiol.* **28**, 229–289 (1965).
- Shatz, C. J. *J. comp. Neurol.* **173**, 497–518 (1977).
- Ferster, D. *J. Neurosci.* **6**, 1284–1301 (1986).
- Douglas, R. J., Martin, K. A. C. & Whittendge, D. *Nature* **332**, 642–644 (1988).
- Ferster, D. & Jagadeesh, B. *J. Neurosci.* **12**, 1262–1274 (1992).
- Nelson, S., Toth, L., Sheth, B. & Sur, M. *Science* **265**, 774–777 (1994).
- Douglas, R. J., Koch, C., Mahowald, M., Martin, K. A. C. & Suarez, H. H. *Science* **269**, 981–985 (1995).
- Saul, A. B. & Humphrey, A. L. *J. Neurophysiol.* **64**, 206–224 (1990).
- Maex, R. thesis, Katholieke Univ. Leuven (1994).
- Suarez, H., Koch, C. & Douglas, R. J. *Neurosci.* **15**, 6700–6719 (1995).
- Tanaka, K. *J. Neurophysiol.* **49**, 1303–1318 (1983).
- Reid, R. C. & Alonso, J.-M. *Nature* **378**, 281–284 (1995).
- Jones, J. P. & Palmer, L. A. *J. Neurophysiol.* **58**, 1187–1211 (1987).
- Chapman, B., Zahs, K. R. & Stryker, M. P. *J. Neurosci.* **11**, 1347–1358 (1991).
- Chapman, B. & Stryker, M. P. *J. Neurosci.* **13**, 5251–5262 (1993).
- Miller, K. D. *J. Neurosci.* **14**, 409–441 (1994).
- Schwartz, H. D., Malpel, J. G., Weyand, T. G. & Lee, C. J. *Neurophysiol.* **56**, 1074–1087 (1986).

ACKNOWLEDGEMENTS. We thank M. Stryker and R. Shapley for suggestions concerning the design and completion of these experiments, and J. Malpel for the use of his cooling apparatus, and for providing invaluable advice on the cooling method.

Functional recovery in parkinsonian monkeys treated with GDNF

Don M. Gash, Zhiming Zhang, Aliza Ovadia, Wayne A. Cass, Ai Yi, Linda Simmerman, Deborah Russell*, David Martin*, Paul A. Lapchak*, Frank Collins*, Barry J. Hoffer† & Greg A. Gerhardt†

Department of Anatomy and Neurobiology, University of Kentucky, College of Medicine, Lexington, Kentucky 40536, USA

* AMGEN Inc., Thousand Oaks, California 91320-1789, USA

† Departments of Pharmacology and Psychiatry, Neuroscience Training Program and Rocky Mountain Center for Sensor Technology, University of Colorado Health Sciences Center, Denver, Colorado 80262, USA

PARKINSON'S disease results from the progressive degeneration of dopamine neurons that innervate the striatum^{1,2}. In rodents, glial-cell-line-derived neurotrophic factor (GDNF) stimulates an increase in midbrain dopamine levels, protects dopamine neurons from some neurotoxins, and maintains injured dopamine neurons^{3–9}. Here we extend the rodent studies to an animal closer to the human in brain organization and function, by evaluating the effects of GDNF injected intracerebrally into rhesus monkeys that have had the symptomatology and pathophysiological features of Parkinson's disease induced by the neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP)^{10–14}. The recipients of GDNF displayed significant improvements in three of the cardinal symptoms of parkinsonism: bradykinesia, rigidity and postural instability. GDNF administered every four weeks maintained functional recovery. On the lesioned side of GDNF-treated animals, dopamine levels in the midbrain and globus pallidus were twice as high, and nigral dopamine neurons were, on average, 20% larger, with an increased fibre density. The results indicate that GDNF may be of benefit in the treatment of Parkinson's disease.

GDNF has been proposed as a potential treatment for parkinsonism on the basis of its effects on rodent dopamine neurons^{3,4,7,8}, but little is known about its actions in non-human primates. Therefore, we have evaluated the effects of GDNF in rhesus monkeys with stable parkinsonian features induced by the infusion of MPTP through the right carotid artery three months earlier^{10–14}. In the first experimental series, sterile MRI-guided stereotaxic procedures¹ were used to deliver a single injection of human recombinant GDNF into the right side of the brain of the rhesus monkeys by one of three routes: intranigral ($n = 2$, 150 µg), intracaudate ($n = 2$, 450 µg) or intracerebroventricular (ICV; $n = 2$, 450 µg). Controls (intranigral, $n = 2$; intracaudate, $n = 2$; or ICV, $n = 3$) received vehicle (10 mM PBS) injections. Using a non-human primate parkinsonian-rating scale¹⁰, parkinsonian features were scored blindly by at least two raters from standardized video tapes made biweekly for the 4-week period before GDNF or vehicle administration and for the 4-week period following treatment. GDNF recipients showed significant functional improvements (Fig. 1a) after all three routes of administration by two weeks post-treatment, which continued for the remainder of the 4-week test period.

Testing was continued on the ICV-treated animals to assess the ability of the animals to respond to repeated dosing and to measure dopamine levels in the brain following GDNF treatment. The sample size was increased by adding one additional animal to the group treated with 450 µg and three treated with 100 µg GDNF ICV. Each of the animals treated with vehicle and trophic factor received three injections into the right lateral ventricle spaced at least 4 weeks apart. In recipients of 100 µg and of 450 µg GDNF, the behavioural effects were very pronounced by the third

TABLE 1 Mesencephalic dopamine neurons

Animal (injection site)	Left side			Right side (MPTP-lesioned)		
	TH+ area (mm ²)	Nigral cell number	Nigral cell size (µm ²)	TH+ area (mm ²)	Nigral cell number	Nigral cell size (µm ²)
Vehicle						
r37 (nigra)	132.8	154,133	379 ± 9	39.8	42,666	321 ± 16
r319 (nigra)	140.9	93,333	360 ± 12	26.6	20,266	299 ± 26
r26 (caudate)	174.0	125,333	405 ± 12	48.2	46,933	353 ± 18
r792 (caudate)	140.3	113,066	366 ± 10	30.2	19,200	353 ± 30
Total ± s.e.m.	147.0 ± 9.2	121,466 ± 12,728	378 ± 10	36.2 ± 4.9	32,266 ± 7,292	332 ± 13
GDNF						
r173 (nigra)	195.5	229,866	480 ± 10	88.6	54,400	435 ± 22
r807 (nigra)	276.0	185,600	476 ± 11	188.4	116,266	416 ± 14
r315 (caudate)	174.8	131,200	472 ± 12	59.2	22,933	406 ± 35
r817 (caudate)	180.5	181,866	366 ± 8	82.5	33,600	342 ± 18
Total ± s.e.m.	206.7 ± 23.5*	182,133 ± 20,176*	448 ± 27*	104.7 ± 28.6*	56,800 ± 20,871	400 ± 20*

GDNF recipients displayed a bilateral increase in TH-positive processes in the ventral tegmental area and substantia nigra. Perikaryal cell size was also bilaterally increased in the substantia nigra. The number of TH-positive neurons was increased in the left nigra in the trophic factor recipients, but was highly variable in the right nigra. * $P \leq 0.05$, GDNF versus vehicle same side, one-tailed t-test. Every sixth 40-µm thick coronal tissue section through the substantia nigra was sampled for evaluation (Bioquant Image Analysis System). This entailed analysis of 14 to 20 slides per animal. The area (number of pixels) of the ventral tegmental area and substantia nigra containing TH-positive immunoreactive staining was quantified. The substantia nigra was defined as consisting of all midbrain TH-positive neurons except those interspersed with the oculomotor nerve rootlets. Using an optical fractionator method for unbiased stereological cell counting²⁰⁻²², the number and perikaryal size of TH-positive neurons in the substantia nigra were estimated. On each section, a 200 µm × 200 µm grid was randomly superimposed with a 150 µm × 150 µm counting chamber placed on each fifth intersection. A 20-µm-deep fraction of the counting chamber was determined by a stage encoder attached to the microscope to measure the Z-axis. All neurons completely within the boundaries of the chamber or crossing the upper or right side of the chamber within the 20-µm depth were counted (up to 370 neurons per side per animal) and their perimeter (minus neurites) measured. With this method, there were too few neurons sampled in the ventral tegmental area to obtain accurate estimates.

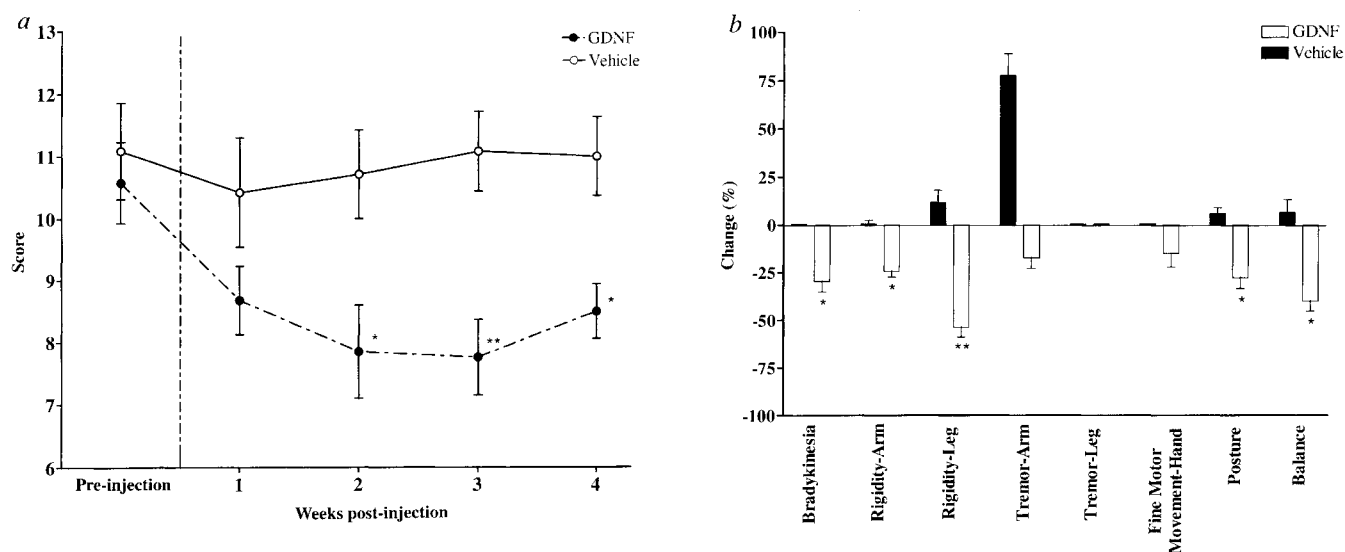


FIG. 1 Behavioural response to GDNF administration. *a*, Behavioural scores for the seven vehicle recipients did not differ from pre-injection ratings at any time point post-injection. However, the GDNF recipient group displayed significantly improved parkinsonian features at weeks 2–4 post-injection. The data are shown as the mean with error bars for the s.e.m. In comparison, parkinsonian rhesus monkeys receiving a single dose of the standard antiparkinsonian drug Sinemet (25 mg carbidopa, 250 mg levodopa) showed improvements from 11 to 9 on the rating scale lasting for several hours¹⁹. Although side effects including dystonias, dyskinesias and vomiting often accompany levodopa treatment in rhesus monkeys, such behaviours were not seen in the GDNF recipients. *b*, The behavioural responses of the ICV recipients during the third week following the third injection are broken out into each of the rated categories. Significant improvements were seen in bradykinesia, rigidity, posture and balance in the GDNF recipients. Upper limb tremor appeared to become more severe in the vehicle-treated animals whereas a tendency for improvement in this feature was seen in the trophic factor recipients. Mean with s.e.m. error bars. * $P \leq 0.05$; ** $P \leq 0.01$; one-tailed Mann–Whitney *U*-test.

METHODS. Adult female rhesus monkeys (Hazelton Research Primates, Alice, TX) with continuously expressed hemiparkinsonian features rated between 8 and 14 on our non-human primate rating scale¹⁰ were randomly assigned to treatment groups. The GDNF (Amgen) doses used were based on effective levels in rodents⁶, scaling up for the larger size of the rhesus monkey brain. Injections of either the vehicle, 10 mM PBS, or GDNF were made using sterile MRI guided stereotaxic procedures¹¹. Intraneural and intracaudate injections were made in 18 µl PBS at the rate of 1 µl min⁻¹; ICV injections were made in 80 µl PBS over 10 min. Parkinsonian features were scored blindly by at least two raters from video tapes of biweekly standardized testing periods. As the rating scale was nonlinear, the non-parametric Mann–Whitney *U*-test was used to analyse differences between groups. Because no differences were seen in the behavioural responses between the routes of administration, all GDNF recipients were treated as one group. Post-injection scores were compared to pre-injection scores (same set of animals).

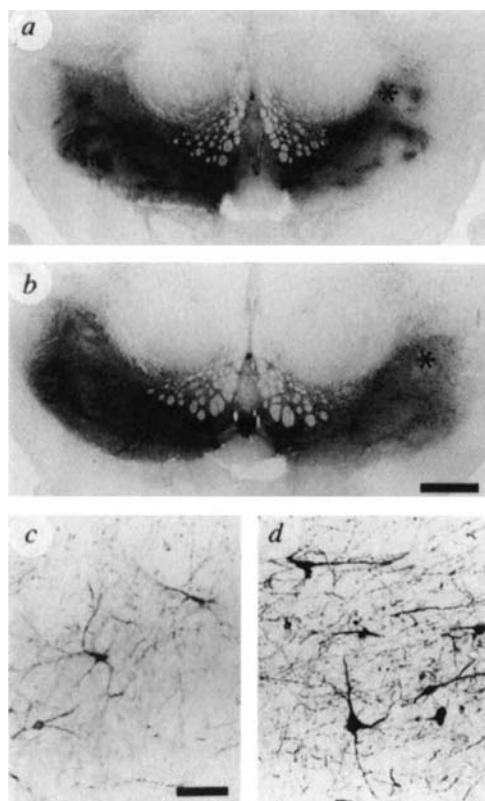


FIG. 2 *a*, As seen in this intranigral vehicle recipient (animal r37), unilateral MPTP administration produces extensive loss of TH-positive staining in dopamine neurons in the lateral substantia nigra on the lesioned side (*) as well as some apparent loss on the contralateral side. *b*, In comparison, quantitative image analysis (Table 1) confirmed that GDNF-recipients such as this animal (r807) displayed bilaterally increased TH-positive fibre staining in the mesencephalic dopaminergic nuclei. Scale bar for *a* and *b*, 2 mm. A higher magnification of the areas in *a* and *b*, demarcated by the asterisk, demonstrates a difference in fibre staining and cell size between a vehicle-treated animal (*c*) and a GDNF recipient (*d*). Scale bar for *c* and *d*, 100 μ m.

METHODS. Four weeks after GDNF or vehicle administration, animals were killed after receiving an overdose of anaesthetic and transcardially perfused with ice-cold saline. The brain was removed and cut into 4-mm-thick sections using a brain mold. The tissue slabs were then immersion-fixed in a 4% paraformaldehyde solution and processed so that 40- μ m-thick coronal sections could be cut on a sliding knife microtome for immunohistochemical staining and image analysis. Immunohistochemical staining for TH followed procedures described previously¹¹.

week following treatment. Significant improvements were evident in bradykinesia, rigidity, posture and balance (Fig. 1*b*). Behavioural side-effects from GDNF injections were limited to a transient post-operative weight loss, which recovered by 4 weeks post-treatment. In vehicle recipients, upper limb tremor increased in severity, whereas the other behavioural features showed no change over the same test period.

The intracaudate- and intranigral-injected animals were killed 4 weeks after the first injection and their midbrains processed for quantitative anatomical analysis. Immunocytochemical staining for tyrosine hydroxylase (TH), the rate-limiting enzyme in dopa-

TABLE 2 Dopamine and dopamine metabolite levels in ICV-treated hemiparkinsonian rhesus monkeys

	DA	DOPAC	HVA	HVA/DA
Substantia nigra				
Vehicle				
left	876 $\pm$ 390	1,441 $\pm$ 809	7,076 $\pm$ 2,394	9.4 $\pm$ 1.9
right	60 $\pm$ 20	129 $\pm$ 42	2,160 $\pm$ 612	42 $\pm$ 13
GDNF				
left	723 $\pm$ 144	1,213 $\pm$ 192	7,803 $\pm$ 650	13 $\pm$ 2.7
right	123 $\pm$ 19*	220 $\pm$ 64	2,852 $\pm$ 195	26 $\pm$ 3.8
Ventral tegmental area				
Vehicle				
left	977 $\pm$ 253	1,812 $\pm$ 356	6,874 $\pm$ 119	7.9 $\pm$ 1.6
right	193 $\pm$ 50	542 $\pm$ 207	3,291 $\pm$ 917	18 $\pm$ 3.4
GDNF				
left	1,218 $\pm$ 162	2,562 $\pm$ 501	7,050 $\pm$ 518	6.7 $\pm$ 1.5
right	407 $\pm$ 63*	829 $\pm$ 167	3,975 $\pm$ 645	9.6 $\pm$ 0.7*
Caudate nucleus				
Vehicle				
left	3,248 $\pm$ 710	5,313 $\pm$ 1,258	11,042 $\pm$ 1,452	3.7 $\pm$ 0.7
right	12 $\pm$ 5	49 $\pm$ 14	652 $\pm$ 244	42 $\pm$ 19
GDNF				
left	2,234 $\pm$ 349	4,960 $\pm$ 646	11,698 $\pm$ 526	6.8 $\pm$ 2.1
right	13 $\pm$ 2	56 $\pm$ 10	740 $\pm$ 44	70 $\pm$ 15
Putamen				
Vehicle				
left	3,269 $\pm$ 449	6,358 $\pm$ 1,851	18,144 $\pm$ 2,609	5.0 $\pm$ 0.5
right	20 $\pm$ 8	81 $\pm$ 27	1,268 $\pm$ 467	46 $\pm$ 9.2
GDNF				
left	2,716 $\pm$ 468	6,640 $\pm$ 598	18,169 $\pm$ 1,577	7.5 $\pm$ 1.1
right	16 $\pm$ 3	66 $\pm$ 11	1,143 $\pm$ 97	83 $\pm$ 15
Nucleus accumbens				
Vehicle				
left	1,821 $\pm$ 305	4,332 $\pm$ 2,105	12,843 $\pm$ 2,986	7.0 $\pm$ 0.6
right	999 $\pm$ 414	1,602 $\pm$ 528	5,229 $\pm$ 1,306	6.2 $\pm$ 1.9
GDNF				
left	1,700 $\pm$ 271	5,540 $\pm$ 789	15,439 $\pm$ 3,940	9.8 $\pm$ 1.1
right	906 $\pm$ 191	2,044 $\pm$ 411	7,549 $\pm$ 742	9.9 $\pm$ 1.9
Globus pallidus				
Vehicle				
left	84 $\pm$ 10	86 $\pm$ 27	5,711 $\pm$ 845	70 $\pm$ 12
right	13 $\pm$ 1	28 $\pm$ 12	961 $\pm$ 129	78 $\pm$ 16
GDNF				
left	93 $\pm$ 11	92 $\pm$ 11	7,348 $\pm$ 687	82 $\pm$ 8.5
right	23 $\pm$ 2*	31 $\pm$ 5	1,420 $\pm$ 125*	63 $\pm$ 2.2

Levels (mean ng g⁻¹ weight $\pm$ s.e.m.) of dopamine (DA), 3,4-dihydroxyphenylacetic acid (DOPAC) and homovanillic acid (HVA) were measured using high-performance liquid chromatography¹¹. The samples are from three control animals which received ICV vehicle injections and six ICV GDNF-treated monkeys (three 100 μ g recipients and three 450 μ g recipients). The animals were killed 3 weeks after the third ICV injection and tissue punches collected using procedures described elsewhere¹¹. The number of punches taken per side per animal were: caudate nucleus, 15; putamen, 10; nucleus accumbens, 5; globus pallidus, 5; ventral tegmental area, 1; and substantia nigra, 3. The values from each site were averaged to obtain a single value per site per animal. Dopamine depletion was extensive on the right side of the brain in all animals from MPTP infused through the right internal carotid artery months earlier. * $P \leq 0.05$ versus vehicle right side, one-tailed t-test.

mine synthesis, was used to identify neuritic processes and cell bodies of surviving dopamine neurons (Fig. 2). There was a distinctive bilateral effect on mesencephalic dopamine neurons in the GDNF recipients with significant increases in fibre staining and cell size (Table 1). More neurons in the substantia nigra tended to display TH-positive staining (the number was significantly higher on the left side, but more variable in the right nigra) suggesting that GDNF upregulated expression of this enzyme in injured dopamine neurons. Three weeks after the third injection, the ICV-treated animals were killed and multiple tissue punches were taken from the basal ganglia and midbrain to measure levels

of dopamine and its metabolites¹¹. The GDNF-induced changes in dopamine levels seen in the rhesus monkeys were limited to the substantia nigra, ventral tegmental area and globus pallidus (Table 2). MPTP-induced depletion of dopamine in the caudate nucleus and putamen was not reversed by GDNF administration.

The behavioural effects seen after all three routes of intracerebral administration suggests that GDNF is able to diffuse effectively through the brain parenchyma to target cells and/or their processes. The bilateral increase in dopamine neuron perikaryal size and TH-positive fibres in the substantia nigra following unilateral injections also supports the widespread distribution of intracerebrally injected GDNF. Evidence for bilateral effects in the substantia nigra from unilateral GDNF injections has also been reported in MPTP-treated mice⁸ and 6-hydroxydopamine lesioned rats^{5,6}. The globus pallidus, substantia nigra and ventral tegmental area all play important roles in regulating motor functions^{15–18}, and it is possible that GDNF's effects on dopamine levels in these nuclei may account for some or all of the observed behavioural improvements. However, future studies are needed to determine whether GDNF induces changes in dopamine regulation (release and uptake) and/or dopamine receptors in the caudate and putamen that contribute to the recovery of movement functions.

Our study extends the findings of potent dopaminergic neurotrophic effects by GDNF reported in cell culture and rodents to the more complex primate brain. The improvements promoted by GDNF in the rhesus monkeys in bradykinesia, rigidity and postural stability directly correspond to behavioural dysfunctions in Parkinson's disease which are difficult to assess in rodents. In addition, the improvement in parkinsonian features, lasting for at least 3 weeks following a single injection of GDNF, was greater than we have seen in parkinsonian monkeys treated with the

standard antiparkinsonian drug, levodopa, in which a single dose only improved parkinsonian features for several hours¹⁹. The direct effects of GDNF on nigral dopamine neurons may account for the behavioural recovery displayed by the parkinsonian monkeys. However, much remains to be learned about the effects of GDNF in the brain, and other neurotransmitter systems and pathways may have been involved. Clearly, additional studies in rodent and primate model systems are warranted to define further the therapeutic potential of GDNF in treating Parkinson's disease. □

Received 30 October 1995; accepted 26 January 1996.

1. Jellinger, K. *Adv. Neurol.* **45**, 1–18 (1986).
2. Hornykiewicz, O. & Kish, S. J. *Adv. Neurol.* **45**, 19–34 (1986).
3. Lin, L.-F. H., Doherty, D. H., Lile, J. D., Bektesh, S. & Collins, F. *Science* **260**, 1130–1132 (1993).
4. Beck, K. D. et al. *Nature* **373**, 339–341 (1995).
5. Bowenkamp, K. E. et al. *J. comp. Neurol.* **355**, 479–489 (1995).
6. Hoffer, B. J. et al. *Neurosci. Lett.* **182**, 107–111 (1994).
7. Hudson, J. et al. *Brain Res. Bull.* **36**, 425–432 (1995).
8. Tomac, A. et al. *Nature* **373**, 335–339 (1995).
9. Kearns, C. M. & Gash, D. M. *Brain Res.* **672**, 104–111 (1995).
10. Ovidia, A., Zhang, Z. & Gash, D. M. *Neurobiol. Aging* **16**, 931–937 (1995).
11. Gash, D. M. et al. *J. comp. Neurol.* **363**, 345–358 (1995).
12. Bankiewicz, K. S. et al. *Life Sci.* **39**, 7–16 (1986).
13. Bankiewicz, K. S., Mandel, R. J. & Sofroniew, M. V. *Expl Neurol.* **124**, 140–149 (1993).
14. Smith, R. D., Zhang, Z., Kurlan, R., McDermott, M. & Gash, D. M. *Neuroscience* **52**, 7–16 (1993).
15. DeLong, M. R. *Trends Neurosci.* **13**, 281–289 (1990).
16. Robertson, G. S. & Robertson, H. A. *Neurosci. Lett.* **89**, 204–208 (1988).
17. Robertson, G. S. & Robertson, H. A. *J. Neurosci.* **9**, 3326–3331 (1989).
18. Dougle, K. L. & Crocker, A. D. *Proc. natn. Acad. Sci. U.S.A.* **92**, 1669–1673 (1995).
19. Greenamyre, J. T. et al. *Ann. Neurol.* **35**, 655–661 (1994).
20. Coggeshall, R. E. *Trends Neurosci.* **15**, 9–13 (1992).
21. West, M. J. *Neurobiol. Aging* **14**, 275–285 (1993).
22. Harding, A. J., Halliday, G. M. & Cullen, K. J. *Neurosci. Meth.* **51**, 83–89 (1994).

ACKNOWLEDGEMENTS. We thank A. Andersen for assistance with the statistical analysis. This work was supported by a contract from Amgen and grants from the USPHS.

Voltage-dependent modulation of N-type calcium channels by G-protein $\beta\gamma$ subunits

Stephen R. Ikeda

Department of Pharmacology and Toxicology, Medical College of Georgia, Augusta, Georgia 30912-2300, USA

THE most commonly used signal transduction pathway for receptor-mediated N-type Ca^{2+} -channel modulation involves activation of a heterotrimeric G protein to produce voltage-dependent inhibition¹. Although it is widely assumed that $G\alpha$ mediates this effect, experiments to address this hypothesis directly are lacking. Here I show that transient overexpression of $G\beta\gamma$ in sympathetic neurons mimics and occludes the voltage-dependent Ca^{2+} channel modulation produced by noradrenaline (NA). Conversely, overexpression of $G\alpha$ produces minimal effects on basal Ca^{2+} channel behaviour but attenuates NA-mediated inhibition in a manner consistent with the buffering of $G\beta\gamma$. These observations indicate that it is $G\beta\gamma$, and not $G\alpha$, that mediates voltage-dependent inhibition of N-type Ca^{2+} channels. The identification of $G\beta\gamma$ as the mediator of this pathway has broad implications as G-protein-coupled receptors, many of which are implicated in disease or are targets of therapeutic agents, couple to N-type Ca^{2+} channels² and may modulate synaptic transmission by this mechanism^{3,4}.

Whole-cell Ca^{2+} currents (I_{Ca}) were recorded from isolated neurons (adult rat sympathetic) that express primarily N-type Ca^{2+} channels^{5,6}. When elicited with a depolarizing voltage pulse, I_{Ca} rose rapidly to a peak and thereafter declined slowly (Fig. 1a, lower trace). In contrast, a biphasic I_{Ca} rising phase, which

has been referred to as 'kinetic slowing'⁷, was evident when G proteins were activated by binding of NA to α_2 -adrenergic receptors⁸ (Fig. 1a, upper trace). Direct G-protein activation by intracellular application of guanylyl-imidodiphosphate (GppNHp), a non-hydrolysable analogue of GTP, also produced a nearly identical kinetic slowing (Fig. 1b). Thus, although it is well established that heterotrimeric G proteins mediate this response, it has not been determined whether it is $G\alpha$ or $G\beta\gamma$ (or both) that interacts with a subsequent element(s) in the transduction pathway¹. To address this question directly, components of heterotrimeric G proteins were overexpressed in sympathetic neurons by intranuclear microinjection of mammalian expression vectors containing complementary DNA inserts coding for molecularly defined G-protein subunits. Figure 1c illustrates the I_{Ca} recorded from a neuron that had been previously injected with both $G\beta_1$ and $G\gamma_2$ (denoted $G\beta_1\gamma_2$) cDNA. In the absence of agonist or intracellular G-protein activator, the I_{Ca} in neurons injected with $G\beta_1\gamma_2$ cDNAs displayed kinetic slowing virtually identical to that produced by application of NA or GppNHp. Similar results were obtained in neurons injected with $G\beta_1\gamma_3$ or $G\beta_1\gamma_7$ cDNAs (results not shown). However, in neurons injected with either $G\beta_1$ or $G\gamma_2$ cDNA alone (Fig. 1d and 1e, respectively), the I_{Ca} rising phase was similar to that of uninjected neurons. These results suggest that the expression and assembly of the $\beta\gamma$ dimer is both required and sufficient to produce Ca^{2+} -channel modulation. In contrast, injection of cDNA coding for a constitutively active $G\alpha_{\text{OA}}$ (Q205L mutant) produced no apparent kinetic slowing (Fig. 1f). The Q205L mutation impairs intrinsic GTPase activity thereby conferring an active phenotype⁹. As G_o has been implicated in coupling of the α_2 -adrenoceptor to I_{Ca} inhibition in rat sympathetic neurons¹⁰, these data suggest that GTP-bound $G\alpha_o$ serves no direct role in mediating voltage-dependent I_{Ca} inhibition.

The ability of $G\beta\gamma$ to mimic closely NA-mediated I_{Ca} inhibition was further examined as illustrated in Fig. 2. Although the exact mechanism responsible for kinetic slowing is unclear, the phen-

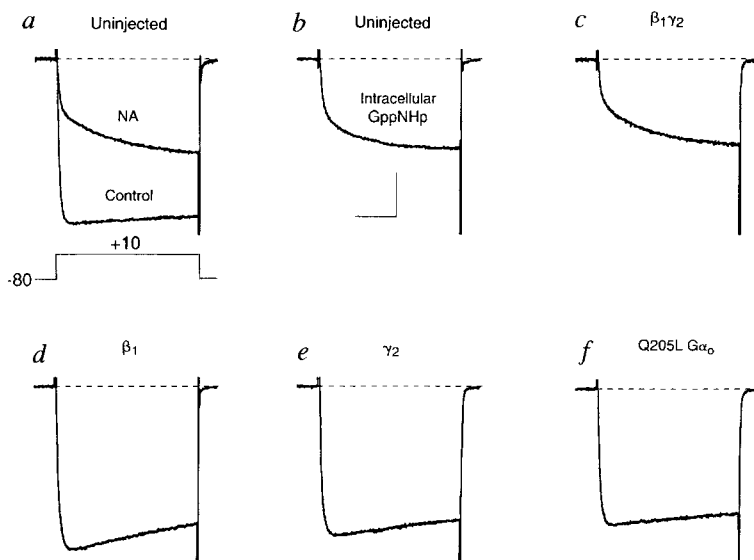
omenon seems to result from a voltage-dependent relief of inhibition that occurs at depolarized potentials^{11–13}. This voltage-dependent effect is clearly demonstrated when a large conditioning depolarization (+80 mV) is inserted between two identical test pulses^{12,13} (Fig. 2a, inset). In the absence of agonist, the I_{Ca} amplitude was little affected by the conditioning pulse (Fig. 2a, lower trace). In contrast, inhibition by NA was greatly relieved by the conditioning pulse (Fig. 2a, upper trace) resulting in an increased ratio of postpulse to prepulse I_{Ca} amplitude termed 'facilitation'. Injection of $G\beta_1\gamma_2$ cDNAs produced significant facilitation in the absence of agonist ('basal facilitation') which mimicked the effect of NA on control neurons (Fig. 2b). Mean basal facilitation was greatly increased in neurons injected with $G\beta_1\gamma_2$, $G\beta_1\gamma_3$ or $G\beta_1\gamma_7$ cDNAs when compared with control neurons (Fig. 2c). Conversely, injection of $G\beta_1$ or $G\gamma_2$ cDNA alone had little effect on mean basal facilitation. Consistent with the idea that overexpression of $G\beta\gamma$ produces nearly maximal tonic modulation, application of NA to neurons injected with $G\beta_1\gamma_2$ cDNAs produced little additional effect (Fig. 2b). Mean inhibition of I_{Ca} amplitude by NA was greatly attenuated in neurons injected with $G\beta_1\gamma_2$, $G\beta_1\gamma_3$ or $G\beta_1\gamma_7$ cDNAs when compared with control neurons or neurons injected with $G\beta_1$ or $G\gamma_2$ cDNA alone (Fig. 2d).

Another strategy to affirm the role of $G\beta\gamma$ in neurotransmitter-mediated I_{Ca} modulation involved increasing the normal $\alpha : \beta\gamma$ ratio by overexpressing $G\alpha_o$. Excess GDP-bound $G\alpha_o$, with a high

affinity for $G\beta\gamma$, should create conditions favouring heterotrimer formation and thus produce a ' $G\beta\gamma$ sink'¹⁴. In accordance with this prediction, injection of cDNA coding for various $G\alpha_o$ subunits significantly attenuated NA-mediated I_{Ca} modulation (Fig. 3a; see Fig. 1a). In neurons injected with Q205L $G\alpha_o$ cDNA, mean NA-mediated I_{Ca} inhibition was decreased by about 60% when compared with uninjected neurons (Fig. 3b). In contrast, injection of wild-type (WT) $G\alpha_o$ or G204A $G\alpha_o$ (a mutant lacking high-affinity GTP- γ S binding but retaining the ability to bind $G\beta\gamma$ ^{15,16}) cDNA nearly abolished NA-mediated I_{Ca} inhibition (Fig. 3b). The ability of Q205L $G\alpha_o$ partly to attenuate I_{Ca} inhibition is consistent with the fact that a cognate mutation in $G\alpha_s$ (Q227L) produces a constitutively active phenotype in which a fraction of the mutant subunits are in the GDP-bound state as indicated by coupling to β -adrenergic receptors¹⁷. Interestingly, the rank order for attenuation of I_{Ca} inhibition paralleled the ability of the different $G\alpha_o$ subunits to remain in the GDP-bound state (that is, G204A > WT > Q205L) as predicted by the $G\beta\gamma$ sink hypothesis. However, as $G\alpha_o$ expression levels were not quantified, interpretations based on the relative magnitude of these effects must be considered with caution. Unlike injection of $G\beta\gamma$ cDNAs (Fig. 2c), injection of $G\alpha_o$ cDNAs did not produce an increase in mean basal facilitation (Fig. 3d). In fact, in neurons injected with WT $G\alpha_o$ or G204A $G\alpha_o$ cDNA, a large depolarizing conditioning pulse produced inactivation (that is, facilitation <1; Fig. 3c, right) rather than the small but consistent basal facilitation seen in

FIG. 1 Kinetic slowing of Ca^{2+} currents (I_{Ca}) in rat sympathetic neurons. I_{Ca} displayed a biphasic rising phase termed 'kinetic slowing' after receptor-mediated G-protein activation with 10 μ M NA (a, upper trace) or direct G-protein activation by intracellular dialysis with 500 μ M GppNHP, a non-hydrolysable GTP analogue (b). Neurons previously injected with $G\beta_1$ and $G\gamma_2$ ($G\beta_1\gamma_2$) cDNA displayed a similar slowing (c) in the absence of agonist or GppNHP. In contrast, neurons previously injected with cDNA coding for $G\beta_1$ alone (d), $G\gamma_2$ alone (e), or a constitutively active $G\alpha_o$, Q205L $G\alpha_o$ (f), exhibited an I_{Ca} resembling that obtained from uninjected neurons in the absence of agonist (a, lower trace). Vertical and horizontal calibrations represent 0.5 nA and 20 ms, respectively.

METHODS. The preparation of isolated adult (150–300 g) rat superior cervical ganglion (SCG) neurons, short-term tissue culture conditions, electrophysiological recording technique, data acquisition and data analysis have been described previously^{28,29}. For recording of I_{Ca} , the external solution consisted of (in mM): tetraethylammonium hydroxide (TEA, 140), HEPES (10), $CaCl_2$ (10), glucose (15), tetrodotoxin (0.0001), pH adjusted to 7.4 with methanesulphonic acid. Patch electrodes (2–6 M Ω) were filled with a solution consisting of (in mM): N-methyl-D-glucamine (125), TEA (20), HEPES (10), EGTA (11), $CaCl_2$ (1), Tris creatine phosphate (14), MgATP (4), Na_2GTP (0.3), pH adjusted to 7.2 with methanesulphonic acid and HCl (final Cl^- concentration of 20 mM). All recordings were made at room temperature (21–24 °C) on neurons maintained in tissue culture for 14–24 h. **Microinjection.** Plasmids coding for G-protein subunits were injected directly into the nucleus of SCG neurons by a variation of the method developed for cytoplasmic injection of RNA²⁹. Briefly, cDNA was diluted with water from stock solutions (about 0.3–1.0 μ g μ l⁻¹) stored in TE buffer (10 mM Tris, 1 mM EDTA, pH 7.4) to a final concentration of 0.1 μ g μ l⁻¹ (per subunit). A 'reporter' plasmid encoding the S65T mutant of the jellyfish green fluorescent protein (GFP)³⁰ was also included (0.1 μ g μ l⁻¹) to facilitate later identification of neurons receiving a successful nuclear injection. After centrifugation (16,000 g for 20 min) to remove particulate matter, the cDNA-containing solution was loaded into a borosilicate micropipette and injected into the nucleus of SCG neurons with an Eppendorf (Madison, WI) 5242 microinjector and 5171 micromanipulator system using injection pressure and duration of 120–200 hPa and 0.3–0.4 s, respectively. Because the isolated neurons were 20–40 μ m in diameter and spherical, the percentage of successful nuclear injections was low (estimated at 10%) and variable, presumably from missing the target in the depth plane. However, successfully injected neurons were readily identified 14–24 h later from the fluorescence produced by the S65T



GFP as observed with an inverted microscope (Diaphot, Nikon, Japan) equipped with an epifluorescence unit (B-2A filter cube, Nikon). The current density (33.0 ± 4 pA pF⁻¹) and magnitude of NA-mediated I_{Ca} inhibition ($59.7 \pm 3.0\%$) in neurons ($n = 9$) injected with S65F GFP cDNA (0.1μ g μ l⁻¹) was not significantly different from that of uninjected neurons (34.8 ± 5 pA/pF and $58.1 \pm 3\%$, respectively, $n = 13$). **Plasmids.** Bovine $G\beta_1$, $G\gamma_2$ and $G\gamma_3$ (M. Simon) and the S65T GFP (R. Heim and R. Tsien) were unidirectionally subcloned into the mammalian expression vector pCI (Promega, Madison, WI) by standard molecular biology techniques and the presence of the insert confirmed by restriction analysis. Restriction sites were added to $G\beta_1$ and S65T GFP by PCR to facilitate subcloning. Rat $G\gamma_7$ was supplied in pCMV5 (L. de Lecea and J. G. Sutcliffe). Rat $G\alpha_o$ subunits (wild-type, Q205L, G204A; S. Strittmatter) cDNAs were supplied in pcDNA1. Plasmids were propagated in either XL-1 Blue or MC1061/p3 *Escherichia coli*, as appropriate, and purified with Qiagen (Chatsworth, CA) midiprep columns. **Drugs.** Noradrenaline HCl (Sigma, St Louis, MO) (10 μ M) was prepared daily from a frozen aqueous stock solution (10 mM). GppNHP (Sigma) was dissolved directly in the pipette solution to a final concentration of 500 μ M on the day of the experiment.

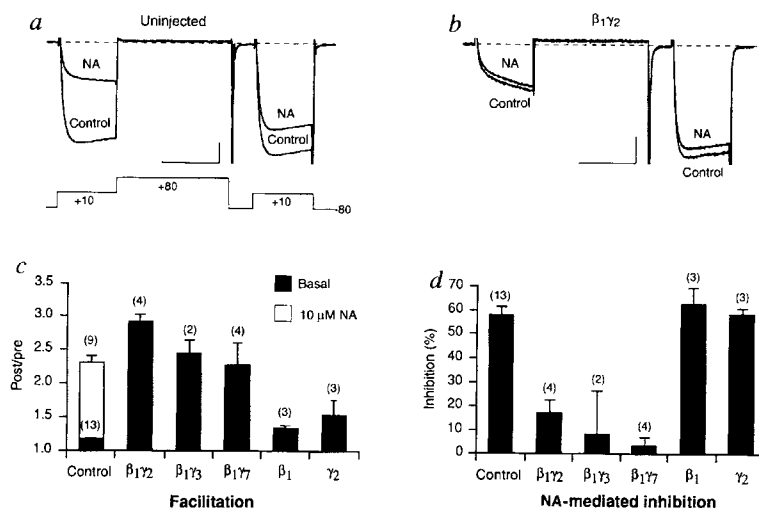
control neurons¹⁸ (Fig. 3c, left)—an effect consistent with the sequestration of G $\beta\gamma$ by overexpressed WT or G204A G α_{oA} .

Finally, if the G $\beta\gamma$ sink hypothesis is correct, it should be possible to overcome the buffering of G $\beta\gamma$ produced by WT G α_{oA} overexpression with intracellular application of GppNHP. Consonant with this idea, the normally rapid onset of facilitation (<4 min to maximum) after dialysis of control neurons (Fig. 4, open circles) with GppNHP was significantly delayed in neurons injected with WT G α_{oA} cDNA (Fig. 4, filled circles). The latter result demonstrated that these neurons were capable of voltage-dependent I_{Ca} inhibition despite the lack of NA-mediated I_{Ca} modulation. The delay in onset of facilitation could be rationalized by assuming that the capacity of overexpressed G α_{oA} to

complex, and thus buffer, G $\beta\gamma$ is eventually depleted as exchange of GDP with GppNHP 'traps' G α_{oA} in an active state with a low affinity for G $\beta\gamma$.

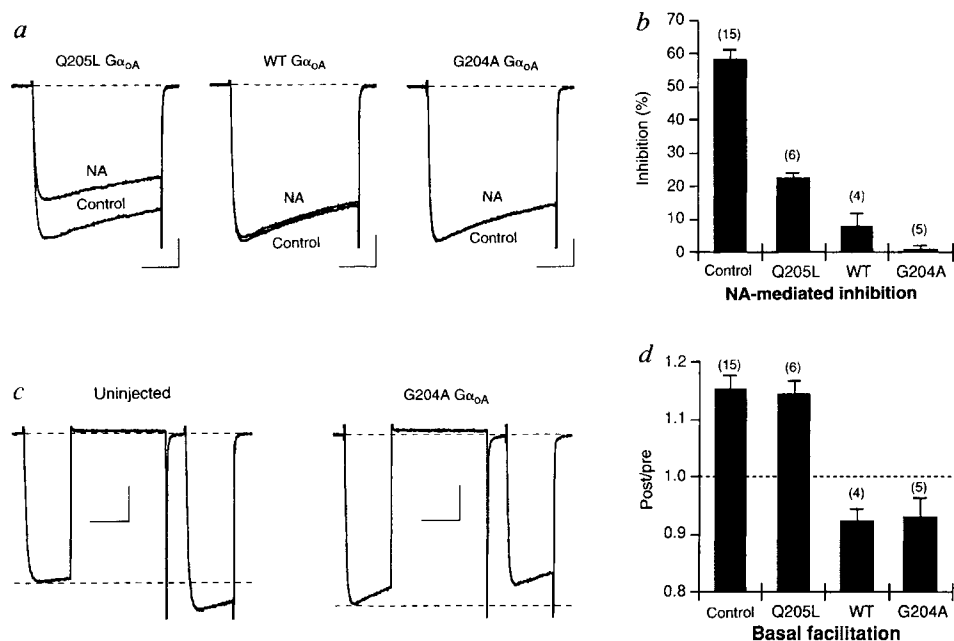
Although a voltage-independent form of N-type Ca^{2+} channel inhibition resulting from G $\beta\gamma$ -mediated activation of a protein-kinase-C-dependent pathway has been reported⁷, these data provide the first evidence that the voltage-dependent inhibition of neuronal N-type Ca^{2+} channels is mediated by G $\beta\gamma$. The data do not support a direct role for G α in mediating voltage-dependent modulation although it is assumed that such elements are essential for receptor–G-protein coupling and termination of G $\beta\gamma$ action. Thus, N-type Ca^{2+} channels join an expanding list of effector molecules that are influenced, possibly directly, by G $\beta\gamma$.

FIG. 2 Tonic modulation in neurons injected with G $\beta\gamma$ cDNA. **a**, Superimposed I_{Ca} traces from a control (uninjected) neuron evoked with the 'double pulse' voltage protocol illustrated (inset) in the absence or presence of 10 μ M NA. **b**, Superimposed I_{Ca} traces in a neuron injected with G $\beta_{1\gamma_2}$ cDNA in the absence or presence of 10 μ M NA. Note the profound kinetic slowing in the prepulse current and large facilitation resulting from the conditioning pulse. Application of 10 μ M NA produced little further modulation. Neurons injected with G $\beta_{1\gamma_3}$ or G $\beta_{1\gamma_7}$ cDNAs displayed similar characteristics (results not shown). Vertical and horizontal calibrations for (a) and (b) are 0.5 nA and 25 ms, respectively. **c**, Summary of mean basal (that is, in the absence of agonist; filled bars) facilitation for control (uninjected) neurons; neurons injected with G $\beta_{1\gamma_2}$, G $\beta_{1\gamma_3}$ or G $\beta_{1\gamma_7}$ cDNAs; and neurons injected with G β_1 or G γ_2 cDNA alone. The open bar represents mean facilitation in uninjected neurons determined during application of 10 μ M NA. Facilitation was calculated as the ratio of I_{Ca} amplitude determined from the test pulse (+10 mV) occurring after (postpulse) and before (prepulse) the +80 mV conditioning pulse. **d**, Summary of mean I_{Ca} amplitude inhibition produced by application of NA to control (uninjected) neurons, neurons injected with G $\beta_{1\gamma_2}$, G $\beta_{1\gamma_3}$ or G $\beta_{1\gamma_7}$ cDNAs and neurons injected with G β_1 or G γ_2 cDNA alone. Inhibition was determined from I_{Ca}



amplitude measured isochronally at 10 ms into the test pulse (+10 mV) in the absence or presence of 10 μ M NA. Data in c and d represent the means $\pm$ s.e.m. for the numbers of neurons indicated in parentheses.

FIG. 3 Attenuation of NA-mediated I_{Ca} inhibition in neurons injected with G α_{oA} cDNA. **a**, Superimposed I_{Ca} current traces recorded from neurons injected with Q205L G α_{oA} (left), wild-type (WT) G α_{oA} (middle) or G204A G α_{oA} (right) cDNA in the absence or presence of 10 μ M NA. The Q205L and G204A mutations produce phenotypes that are constitutively active (decreased GTPase activity) and deficient in high-affinity GTP binding, respectively. Currents were elicited with a 70-ms test pulse to +10 mV from a holding potential of -80 mV. Horizontal calibration is 20 ms. Vertical calibration is 0.25, 0.5 and 0.25 nA, respectively (left to right). **b**, Summary of mean I_{Ca} amplitude inhibition produced by application of NA to neurons injected with Q205L, WT and G204A G α_{oA} cDNA. Data represent the means $\pm$ s.e.m. for the numbers of neurons indicated in parentheses. **c**, I_{Ca} traces, evoked with a 'double pulse' voltage protocol as in Fig. 2a, for an uninjected neuron (left) or neuron injected with G204A G α_{oA} cDNA (right). Horizontal calibration is 20 ms. Vertical calibration is 0.5 and 0.25 nA, respectively (left to right). **d**, Summary of mean basal facilitation for control (uninjected) neurons and



neurons injected with Q205L, WT and G204A G α_{oA} cDNA. Facilitation was calculated as in Fig. 2c. Data in b and d represent the means $\pm$ s.e.m. for the numbers of neurons indicated in parentheses.

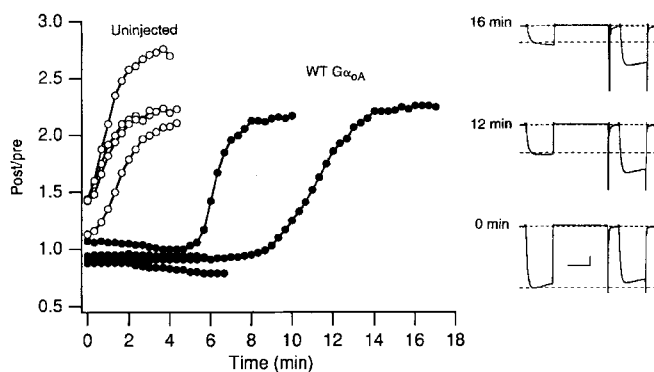


FIG. 4 Time course of guanylyl-imidodiphosphate (GppNHP)-induced I_{Ca} facilitation. Four uninjected (open circles) and four neurons injected with WT $G\alpha_{OA}$ cDNA (filled circles) were dialysed with 500 μ M GppNHP (included in the patch pipette) and facilitation was determined as in Fig. 2c at 20 s intervals. Time zero represents the earliest possible recording after patch rupture (usually about 1 min). Two neurons injected with $G\alpha_{OA}$ cDNA were lost at about 6 min into the recording. Uncompensated series resistance was 4.9 ± 0.4 and 5.2 ± 0.6 M Ω for control neurons ($n = 4$) and neurons injected with $G\alpha_{OA}$ cDNA ($n = 4$), respectively. Current traces were taken from the neuron represented by the rightmost plotted data at the time points indicated. Horizontal and vertical calibrations are 0.5 nA and 25 ms, respectively.

Currently this list includes β -adrenergic receptor kinase, certain forms of adenylyl cyclase, phospholipase C β , and a G-protein-modulated inwardly rectifying K $^{+}$ channel (reviewed in refs 19, 20). Although the precise molecular mechanism by which G $\beta\gamma$ produces voltage-dependent inhibition of N-type Ca $^{2+}$ channels remain to be elucidated, it is intriguing that a G $\beta\gamma$ binding motif, QXXER, identified in the aforementioned proteins²¹, is found within a region on the α_{1B} subunit of the N-type Ca $^{2+}$ channel that interacts with the Ca $^{2+}$ -channel β subunit²². In this regard, several recent studies have demonstrated an interaction between Ca $^{2+}$ -channel β subunits and G-protein modulation^{23–25}. Thus one can speculate that voltage-dependent N-type Ca $^{2+}$ -channel modulation might occur through binding of G $\beta\gamma$ to the Ca $^{2+}$ -channel α_{1B} subunit at a site that alters or disrupts α_{1B} / β -subunit interaction.

The identification of the role that G $\beta\gamma$ plays in N-type Ca $^{2+}$ -channel modulation has broad implications because a great number of G-protein-coupled receptors have been shown to couple to N-type Ca $^{2+}$ channels^{1,2}. Moreover, another Ca $^{2+}$ -channel type implicated in control of synaptic transmission, the P-type, displays a nearly identical form of voltage-dependent modulation²⁶ and thus the findings described here are possibly applicable to this channel type as well. It now appears that ion-channel modulation by G $\beta\gamma$, a controversial finding when first described for an inwardly rectifying K $^{+}$ channel²⁷, could be more common than previously realized. □

Received 21 November 1995; accepted 15 February 1996.

- Hille, B. *Trends Neurosci.* **17**, 531–536 (1994).
- Anwyl, R. *Brain Res. Rev.* **16**, 265–281 (1991).
- Miller, R. *FASEB J.* **4**, 3291–3299 (1990).
- Dunlap, K., Luebke, J. & Turner, T. *Trends Neurosci.* **18**, 89–98 (1995).
- Plummer, M. R., Rittenhouse, A., Kanevsky, M. & Hess, P. *J. Neurosci.* **11**, 2339–2348 (1991).
- Regan, L. J., Sah, D. W. Y. & Bean, B. P. *Neuron* **6**, 269–280 (1991).
- Diversé-Pierluissi, M., Goldsmith, P. K. & Dunlap, K. *Neuron* **14**, 191–200 (1995).
- Schofield, G. G. *Eur. J. Pharmacol.* **180**, 37–47 (1990).
- Strittmatter, S. M., Fishman, M. C. & Zhu, X.-P. *J. Neurosci.* **14**, 2327–2338 (1994).
- Caulfield, M. P. *et al. J. Physiol., Lond.* **477**, 415–422 (1994).
- Bean, B. *Nature* **340**, 153–156 (1989).
- Grassi, F. & Lux, H. D. *Neurosci. Lett.* **105**, 113–119 (1989).
- Elmslie, K. S., Zhou, W. & Jones, S. W. *Neuron* **5**, 75–80 (1990).
- Slepak, V. Z., Katz, A. & Simon, M. I. *J. Biol. Chem.* **270**, 4037–4041 (1995).
- Slepak, V. Z., Wilkie, T. M. & Simon, M. I. *J. Biol. Chem.* **268**, 1414–1423 (1993).
- Slepak, V. Z. *et al. J. Biol. Chem.* **268**, 21889–21894 (1993).
- Masters, S. B. *et al. J. Biol. Chem.* **264**, 15467–15474 (1989).
- Ikeda, S. R. *J. Physiol., Lond.* **439**, 181–214 (1991).

- Clapham, D. E. & Neer, E. J. *Nature* **365**, 403–406 (1993).
- Müller, S. & Lohse, M. J. *Biochem. Soc. Trans.* **23**, 141–148 (1995).
- Chen, H. *et al. Science* **268**, 1166–1169 (1995).
- Pragnell, M. *et al. Nature* **368**, 67–70 (1994).
- Campbell, V., Berrow, N. S., Fitzgerald, E. M., Brickley, K. & Dolphin, A. C. *J. Physiol., Lond.* **485**, 365–372 (1995).
- Roche, J. P., Anantharam, V. & Treisman, S. N. *FEBS Lett.* **371**, 43–46 (1995).
- Bourinet, E. *et al. Soc. Neurosci. Abstr.* **21**, 515 (1995).
- Mintz, I. & Bean, B. P. *Neuron* **10**, 889–898 (1993).
- Logothetis, D. E., Kurachi, Y., Galper, J., Neer, E. J. & Clapham, D. E. *Nature* **325**, 321–326 (1987).
- Zhu, Y. & Ikeda, S. R. *Neuron* **13**, 657–669 (1994).
- Ikeda, S. R., Lovinger, D. M., McCool, B. A. & Lewis, D. L. *Neuron* **14**, 1029–1038 (1995).
- Heim, R., Cubitt, A. B. & Tsien, R. Y. *Nature* **373**, 663–664 (1995).

ACKNOWLEDGEMENTS. I thank E. Zboran for technical assistance, B. Hollins for assistance with the subcloning, and K. Dunlap, D. M. Lovinger and G. G. Schofield for reading an earlier version of the manuscript and providing critical comments; and the following for providing cDNA clones that made this study possible: R. Heim and R. Tsien, S65T GFP; S. Strittmatter, WT $G\alpha_{OA}$, Q205L $G\alpha_{OA}$ and G204A $G\alpha_{OA}$; M. Simon, G β_1 , G β_2 and G β_3 ; L. de Lecea and J. G. Sutcliffe, G γ_1 . This work was supported by a grant from the American Heart Association—Georgia Affiliate.

Modulation of Ca $^{2+}$ channels by G-protein $\beta\gamma$ subunits

Stefan Herlitze*, David E. Garcia†‡, Ken Mackie†§, Bertil Hille†, Todd Scheuer* & William A. Catterall*

Departments of * Pharmacology, † Physiology and Biophysics, and § Anesthesiology, University of Washington, Seattle, Washington 98195, USA

CALCIUM ions entering cells through voltage-gated Ca $^{2+}$ channels initiate rapid release of neurotransmitters and secretion of hormones. Ca $^{2+}$ currents can be inhibited in many cell types by neurotransmitters acting through G proteins via a membrane-delimited pathway independently of soluble intracellular messengers^{1–4}. Inhibition is typically caused by a positive shift in the voltage dependence and a slowing of channel activation and is relieved by strong depolarization resulting in facilitation of Ca $^{2+}$ currents^{1,4–6}. This pathway regulates the activity of N-type and P/Q-type Ca $^{2+}$ channels^{1,2,7}, which are localized in presynaptic terminals^{8,9} and participate in neurotransmitter release^{10–13}. Synaptic transmission is inhibited by neurotransmitters through this mechanism^{1,4}. G-protein α subunits confer specificity in receptor coupling^{1–4,14–17}, but it is not known whether the $G\alpha$ or $G\beta\gamma$ subunits are responsible for modulation of Ca $^{2+}$ channels. Here we report that G $\beta\gamma$ subunits can modulate Ca $^{2+}$ channels. Transfection of G $\beta\gamma$ into cells expressing P/Q-type Ca $^{2+}$ channels induces modulation like that caused by activation of G protein-coupled receptors, but $G\alpha$ subunits do not. Similarly, injection or expression of G $\beta\gamma$ subunits in sympathetic ganglion neurons induces facilitation and occludes modulation of N-type channels by noradrenaline, but $G\alpha$ subunits do not. In both cases, the $G\gamma$ subunit is ineffective by itself, but overexpression of exogenous G β subunits is sufficient to cause channel modulation.

To study the interaction between G-protein subunits and Ca $^{2+}$ channels, we co-transfected the α_{1A} , β_{1b} and $\alpha_2\delta$ subunits^{18,19} of P/Q-type Ca $^{2+}$ channels in tsA-201 cells and analysed the voltage dependence of activation and the facilitation of the Ba $^{2+}$ currents (I_{Ba}) elicited by voltage step protocols (Fig. 1). Activation of endogenous G proteins by intracellular GTP- γ S shifted the mid-point of voltage-dependent activation of I_{Ba} to more positive potentials and made the voltage dependence of activation less steep than recordings without GTP- γ S (Fig. 1a). Facilitation was analysed by comparing I_{Ba} during test pulses before and after a large conditioning prepulse to positive potentials. No facilitation of I_{Ba} in the second test pulse was observed with the control pipette solution (Fig. 1b), but currents could be accelerated and facilitated by prepulses over a wide voltage range in the presence of

‡ Permanent address: Department of Physiology, Faculty of Medicine, UNAM, CP04510 Mexico City D.F., Mexico.

GTP- γ S (Fig. 1d). No differences in voltage dependence or facilitation could be detected with GDP- β S added to the control intracellular solution to inhibit G-protein activation (Fig. 1a, c), indicating that the level of endogenous G-protein modulation in the absence of a stimulus is small in these cells. These results show that transfected P/Q-type Ca^{2+} channels are modulated by GTP- γ S in the same manner that endogenous N-type and P/Q-type Ca^{2+} channels are modulated by G-protein-coupled receptors in neurons^{1,2}.

Membrane-delimited modulation of Ca^{2+} channels most often involves the pertussis-toxin-sensitive G proteins G_o and G_i ¹⁻⁷. To analyse which G-protein subunits are involved in the modulation of transfected P/Q-type Ca^{2+} channels, we co-transfected the pertussis-toxin-sensitive $G\alpha$ subunits $G\alpha_{i1}$, $G\alpha_{i2}$, $G\alpha_{i3}$ or $G\alpha_o$ and the $G\beta\gamma$ subunits $G\beta_2$ and $G\gamma_3$. Co-transfection of $G\beta_2$ and $G\gamma_3$ shifted the voltage dependence of activation to more positive potentials and reduced the steepness of voltage-dependent activation of Ca^{2+} channels (Fig. 2a). I_{Ba} could also be accelerated and facilitated by positive prepulses over the whole voltage range (Fig. 2d). In contrast, only small differences were seen in the voltage dependence of channel activation for $G\alpha_{i1}$, $G\alpha_{i2}$, $G\alpha_{i3}$ and $G\alpha_o$ (Fig. 2a), and no facilitation could be detected (Fig. 2b, c). Nevertheless, coexpression of the $G\alpha$ subunits did significantly reduce facilitation caused by GTP- γ S (Fig. 2, legend), confirming that they were functionally expressed. These results provide direct

evidence that $G\beta\gamma$ subunits, but not $G\alpha$ subunits, can mimic the modulation of Ca^{2+} channels by G-protein-coupled receptors.

G-protein β and γ subunits are thought to be tightly associated *in vivo* and their regulatory effects are ascribed to the intact dimer^{20,21}. Surprisingly, we found that transfection of $G\beta_2$ alone was nearly as effective as $G\beta_2\gamma_3$ in shifting the voltage dependence of Ca^{2+} -channel activation to more positive voltages (Fig. 3a) and in inducing the facilitation and acceleration of I_{Ba} by positive prepulses (Fig. 3c). In contrast, transfection of $G\gamma_3$ alone did not shift the activation curve or induce facilitation (Fig. 3b, d), although the level of expression of $G\gamma_3$ was sufficient to significantly reduce facilitation caused by GTP- γ S (Fig. 3, legend). In view of the tight association of $G\beta\gamma$ ^{20,21}, overexpression of $G\beta$ might lead to increased formation of $G\beta\gamma$ dimers, which modulate Ca^{2+} channels. Alternatively, it is conceivable that $G\beta$ subunits could modulate Ca^{2+} channels directly. In either case, our results suggest that increased expression of endogenous $G\beta$ subunits in an excitable cell would be sufficient for regulation of Ca^{2+} channels.

Endogenous N-type Ca^{2+} channels in superior cervical ganglion (SCG) neurons are modulated by noradrenaline (NA) and other neurotransmitters, which act through multiple G proteins to inhibit Ca^{2+} current¹. In one major pathway of modulation, PTX-sensitive G proteins act through a membrane-delimited mechanism to reduce I_{Ca} in a voltage-dependent manner and

FIG. 1 Effects of guanine nucleotides on the modulation of P/Q-type channel in tsA-201 cells. a, Effects of GDP- β S (2.0 mM) and GTP- γ S (0.6 mM) on the voltage dependence of activation in comparison with the control intracellular solution. Tail currents with Ba^{2+} as current carrier were recorded for 4 ms at a holding potential of -60 mV after a test pulse to the indicated test potentials for 4 ms. Tail currents were normalized to the largest tail current in each series of test pulses, and means $\pm$ s.e.m. were plotted as a function of the test voltage. Control intracellular solution, filled circles; GDP- β S, open squares; GTP- γ S, open triangles. b–d, Facilitation of tail currents with no additions to the intracellular solution (b), 2.0 mM GDP- β S (c), 0.6 mM GTP- γ S (d) after a 10-ms conditioning prepulse to $+100$ mV. For measuring facilitation, a 4-ms test pulse (test 1) to the indicated test potentials was applied from the holding potential of -60 mV. After 1 s, a 10-ms conditioning prepulse to $+100$ mV was applied, the cell was repolarized to -60 mV for 10 ms, and a second 4-ms test pulse (test 2) to the indicated potential was applied. Test pulse 1 and test pulse 2 always stepped to the same test potential. Tail currents were normalized to the largest tail current in each series of test pulses, and means $\pm$ s.e.m. were plotted against test pulse potential. Test 1, open circles; test 2, filled circles. An increase in tail current after test 2 compared with test 1 indicates facilitation. The current traces shown in each panel were recorded during and after test 1 and test 2 at $+25$ mV. METHODS. cDNAs encoding Ca^{2+} channel subunits α_{1A} and β_{1B} (ref. 18) were cloned in pMT2XS, $\alpha_{2\delta}$ (ref. 19) in pZEM228 and CD8 in EBP-pcD. tsA-201 cells were transfected with these cDNAs in 1:1:1 molar ratios plus CD8 in either calcium phosphate or Lipofectamine (Stratagene), incubated for at least 72 h, identified by labelling with anti-CD8 antibody tagged with fluorophore, and analysed by whole-cell patch clamp as described previously²⁸. Currents were recorded and filtered at 10-kHz with an eight-pole Bessel filter. Leak and capacitive currents were subtracted by using hyperpolarizing pulses with the p/4 method. Cells were bathed in an external solution containing 100 mM Tris, 4 mM MgCl_2 , 10 mM BaCl_2 with pH adjusted to 7.3 with methanesulphonic acid. The internal pipette solution consisted of 120 mM aspartic acid, 5 mM CaCl_2 , 2 mM MgCl_2 , 10 mM HEPES, 10 mM EGTA and 2 mM Mg-ATP with pH adjusted to 7.3 with CsOH. Where indicated, GTP- γ S and GDP- β S were added to the internal solution at concentrations of 0.6 and 2.0 mM, respectively.

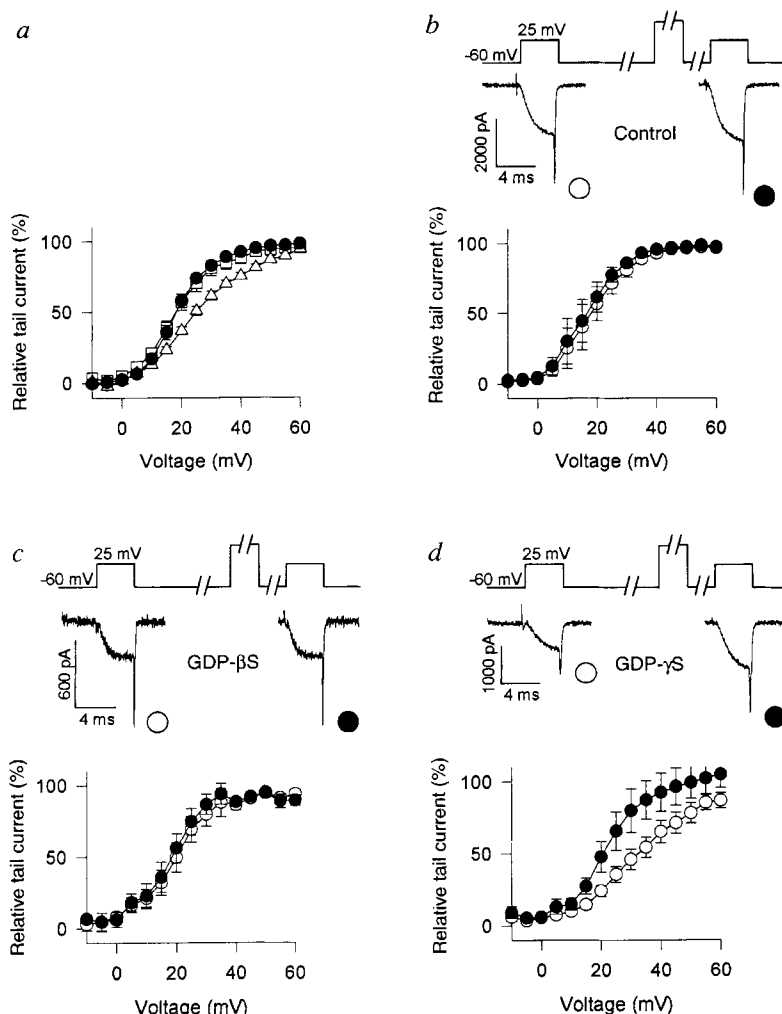


FIG. 2 Effects of $G\alpha$ and $G\beta\gamma$ subunits on voltage dependence of activation and prepulse facilitation of P/Q-type channels in tsA-201 cells. *a*, Voltage dependence of activation of I_{Ba} currents through P/Q-type Ca^{2+} channels transfected alone (filled circles) or with $G\alpha_{oA}$ (open circles), $G\alpha_{i1}$ (open squares), $G\alpha_{i2}$ (open triangles), $G\alpha_{i3}$ (open diamonds), or $G\beta_2\gamma_3$ (filled triangles). Tail currents were recorded and analysed as described in Fig. 1*a*. *b–d*, Effect of $G\alpha_{i1}$ (*b*), $G\alpha_{oA}$ (*c*) and $G\beta_2\gamma_3$ (*d*) on facilitation of the Ba^{2+} currents by a 10-ms prepulse to +100 mV measured and plotted as described in Fig. 1*b–d*. Test 1, open circles; test 2, filled circles. In parallel experiments, we measured facilitation of I_{Ba} in the presence of 0.6 mM GTP- γ S. I_{Ba} was recorded during a 10-ms test pulse (test 1) to +30 mV from a holding potential of -60 mV. The cells were depolarized to +100 mV for 10 ms or 100 ms and a second identical test pulse (test 2) was applied. The facilitation ratio was calculated as the ratio of I_{Ba} at the end of test 2 divided by I_{Ba} at the end of test 1. Average peak I_{Ba} was increased 1.9- to 2.9-fold by coexpression of $G\alpha_{i1}$, $G\alpha_{i2}$, $G\alpha_{i3}$ or $G\alpha_{oA}$, but these increases did not reach statistical significance because of the large variation in peak I_{Ba} from cell to cell. However, coexpression of $G\alpha$ subunits significantly reduced facilitation ratio. For 10-ms conditioning pulses, the control facilitation ratio was 1.97 ± 0.08 (s.e.m., $n = 32$), and it was reduced to 1.42 ± 0.22 (s.e.m., $n = 5$, $P < 0.05$) by coexpression of $G\alpha_{i1}$, to 1.36 ± 0.17 (s.e.m., $n = 3$, $P < 0.05$) by $G\alpha_{i2}$, and to 0.95 ± 0.04 (s.e.m., $n = 6$, $P < 0.01$) for $G\alpha_{oA}$. For 100-ms conditioning pulses, the control facilitation ratio was 1.38 ± 0.10 (s.e.m., $n = 5$), and it was reduced to 0.83 ± 0.23 (s.e.m., $n = 4$, $P = 0.05$) by $G\alpha_{i1}$ and to 0.86 ± 0.11 (s.e.m., $n = 3$, $P < 0.05$) by $G\alpha_{i3}$. The physiological significance of these effects of $G\alpha$ subunits is unknown because they do not mimic known effects of activation of G-protein-coupled receptors. However, they confirm that the $G\alpha$ subunits are effectively expressed in tsA-201 cells under our experimental conditions. **METHODS.** cDNAs encoding $G\alpha_{i1}$ and $G\alpha_{oA}$ were subcloned in pCD-PS, $G\alpha_{i2}$ and $G\alpha_{i3}$ in pNUT, $G\beta_2$ and $G\gamma_3$ in pcDNA-1. These cDNAs were co-transfected with the cDNAs encoding the Ca^{2+} -channel subunits at a ratio of 1:1.

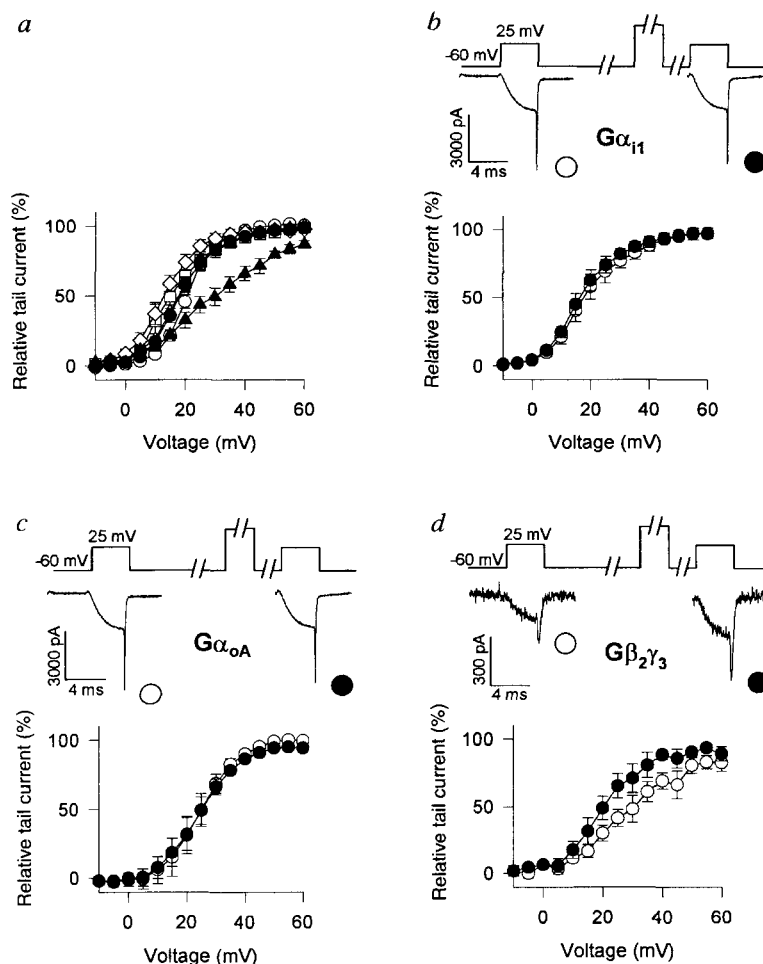
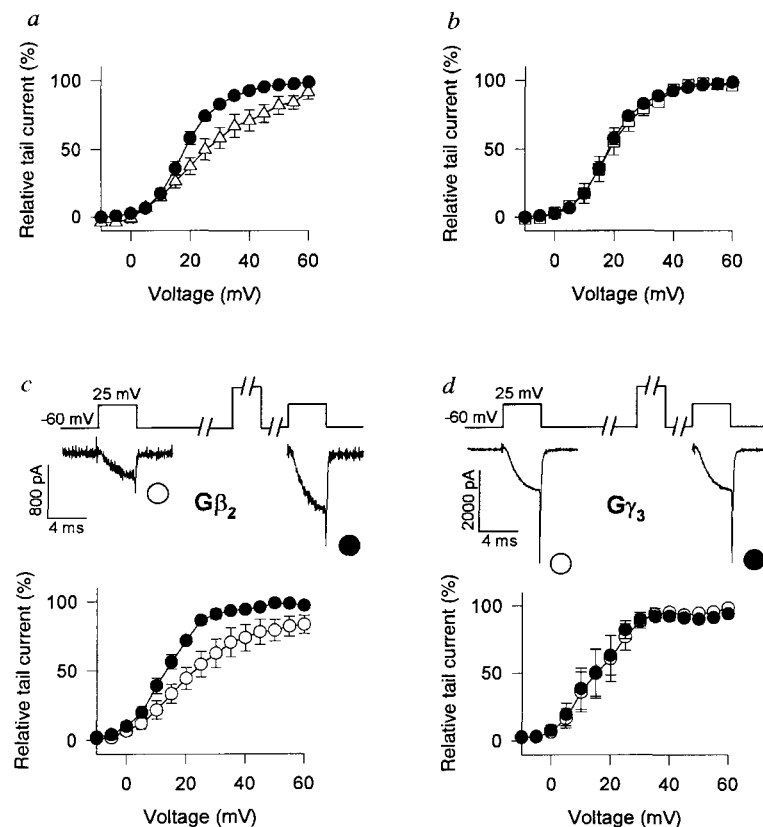


FIG. 3 Effects of $G\beta_2$ and $G\gamma_3$ on voltage dependence of activation and prepulse facilitation of P/Q-type channels in tsA-201 cells. *a, b*, Voltage dependence of activation. tsA-201 cells were transfected with cDNAs encoding Ca^{2+} -channel subunits and no additions (filled circles), $G\beta_2$ (*a*, open triangles), or $G\gamma_3$ (*b*, open squares). Voltage dependence of activation of I_{Ba} was measured and plotted as in Fig. 1*a*. *c, d*, For cells transfected as in *a* and *b*, facilitation of I_{Ba} by 10-ms prepulses to +100 mV was measured and plotted as in Fig. 1*b–d*. Test 1, open circles; test 2, filled circles. In parallel experiments, we measured the facilitation ratio of I_{Ba} in the presence of 0.6 mM GTP- γ S as described in the legend to Fig. 2. For a 10-ms conditioning pulse, the facilitation ratio was reduced from 1.97 for the control to 1.03 ± 0.06 (s.e.m., $n = 3$, $P < 0.01$) for $G\gamma_3$. As for the $G\alpha$ subunits (Fig. 2, legend), these results confirm that the $G\gamma_3$ subunit was effectively expressed in our experiments.



thereby induce facilitation by positive prepulses¹. We injected SCG neurons with purified G $\beta\gamma$ protein or with RNA encoding G α , G β and G γ subunits and determined whether these injections induced tonic modulation of I_{Ca} or occluded the modulatory action of NA. Tonic inhibition of the current was increased in cells injected with G $\beta\gamma$ protein, as judged by a 1.6-fold greater facilitation of I_{Ca} by a depolarizing prepulse in the absence of NA (Fig. 4a). Furthermore, inhibition of I_{Ca} by 10 μ M NA was decreased from 63% to 43% (Fig. 4a). Similarly, in comparison with cells injected with control RNA, cells injected with RNA encoding G $\beta_2\gamma_3$ showed both an increase in tonic inhibition, as indicated by an increase in the prepulse facilitation ratio from 1.2-fold to 2.1-fold, and a decrease in the inhibition by 10 μ M NA from 60% to 32% (Fig. 4b). Injection of RNA encoding G β_2 alone gave comparable or stronger effects; the prepulse facilitation ratio was increased to 4.0-fold and the inhibition by 10 μ M NA was decreased to 36% (Fig. 4b). In contrast, injection of RNA encoding G γ_3 , G α_{oA} or G α_{i3} had no significant effect on inhibition by NA or on prepulse facilitation (Fig. 4b). However, expression of G α_{oA} significantly increased I_{Ca} during control depolarizations, as would be expected if it associated with endogenous G $\beta\gamma$ subunits and thereby relieved tonic inhibition of I_{Ca} (current density 11.1 ± 1.4 pA/pF, $n = 10$ with G α_{oA} , compared with 6.0 ± 0.8 pA/pF, $n = 8$ for controls). Expression of G α_{i3} and G γ_3 did not have significant effects on current density, but did significantly reduce voltage-independent modulation¹ of I_{Ca}

(Fig. 4, legend). Overall, these results indicate that G $\beta\gamma$ subunits play the predominant role in the modulation of endogenous N-type Ca^{2+} channels by NA in SCG neurons, as we have observed for the modulation of P/Q-type Ca^{2+} channels expressed in tsA-201 cells by GTP- γ S. A different, voltage-independent pathway of Ca^{2+} channel modulation involving an intracellular messenger in chick sympathetic neurons might also use $\beta\gamma$ subunits as mediators¹⁷.

G proteins regulate the activity of multiple effector proteins through the action of both G α and G $\beta\gamma$ subunits^{20,21}. Inwardly rectifying K^+ channels are also modulated by G proteins by a membrane-delimited mechanism^{22,23}, and recent evidence suggests that direct binding of G-protein $\beta\gamma$ subunits is involved in their activation²⁴⁻²⁷. We have shown that G $\beta\gamma$ subunits, but not pertussis-toxin-sensitive G α subunits, shift the voltage dependence of activation of the P/Q-type Ca^{2+} channel and induce facilitation of I_{Ca} in response to positive prepulses when coexpressed with the transfected P/Q-type channels or endogenous N-type channels. Because G $\beta\gamma$ subunits cause the same voltage-dependent effects as activation of a membrane-delimited modulatory pathway, the simplest mechanism to explain our results is direct binding to a recognition site on the Ca^{2+} channels, but our data do not exclude more complex regulatory pathways.

Many neurotransmitters acting through G proteins containing different G α subunits cause similar modulation of Ca^{2+} channels¹⁻⁴. Because activation of all G proteins is thought to

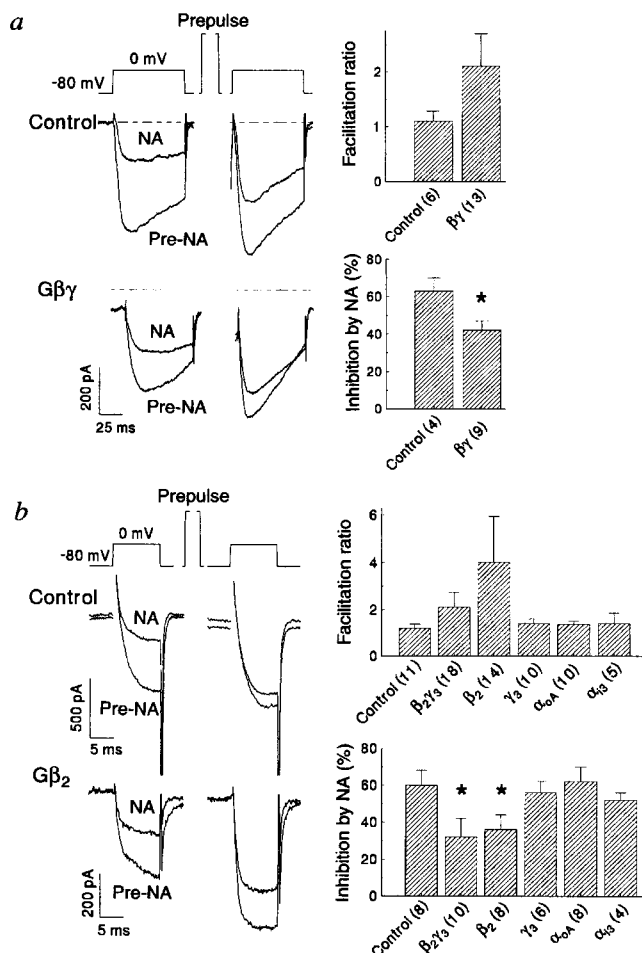


FIG. 4 Enhanced tonic inhibition and decreased modulation of I_{Ca} in SCG neurons mediated by G $\beta\gamma$ subunits. **a**, Injection of G $\beta\gamma$ protein into rat SCG neurons enhances tonic inhibition and decreases modulation of I_{Ca} by 10 μ M NA. Left, current traces from individual cells injected with buffer alone (top) or 2 μ M $\beta\gamma$ protein (bottom). Right, aggregate data demonstrating an increase in tonic inhibition as measured by the facilitation ratio (top) and a decrease in modulation by NA (bottom). **b**, Injection of G $\beta_2\gamma_3$ or β_2 RNA into rat SCG neurons enhances tonic inhibition and decreases modulation of I_{Ca} by 10 μ M NA whereas other subunits are ineffective. Left, current traces from individual cells injected with RNA for cannabinoid CB1 receptor (control, top) or G β_2 (bottom). Right, aggregate data demonstrating that G $\beta_2\gamma_3$ and G β_2 RNA injection increase tonic inhibition as measured by the facilitation ratio (top) and decrease inhibition by NA (bottom), whereas G γ_3 , G α_{oA} , and G α_{i3} cRNA-injected neurons were indistinguishable from control CB1-injected neurons. Facilitation data are presented as means $\pm$ 95% confidence intervals; inhibition data are presented as means $\pm$ s.e.m. Numbers in parentheses are the numbers of cells tested for that condition. *, $P < 0.05$, Mann-Whitney test. In parallel experiments, we also measured the voltage-independent modulation¹ of I_{Ca} by 10 μ M NA. It was $38 \pm 3\%$ for control ($n = 8$) compared with $29 \pm 2\%$ for G α_{i3} ($n = 4$, $P < 0.05$, Mann-Whitney test) or $28 \pm 4\%$ for G γ_3 ($n = 6$, $P < 0.05$, Mann-Whitney test), confirming that these subunits were effectively expressed in SCG neurons.

METHODS. SCG neurons were isolated²⁸ from 5–14-day-old rat pups and cultured³⁰ on laminin and polylysine-coated glass coverslips. For G $\beta\gamma$ protein injections, cells were grown for 12–36 h and injected with purified bovine brain G $\beta\gamma$ in 0.1% fluorescein-coupled Dextran by using an Eppendorf Transjector. After a 30–60-min incubation, fluorescein-labelled cells were selected for recording. For RNA synthesis, G α_{i3} was subcloned into pGEM2 whereas G α_{oA} and G γ were transcribed from pcDNA3. Capped transcripts were prepared from linearized plasmid with a commercial kit (Ambion) at a final concentration of $1\text{--}1.5$ mg ml⁻¹. Eight hours after plating, SCG neurons were injected with RNA as above. After a 12–24-h incubation, recordings were made from fluorescein-labelled neurons. Electrophysiological recordings of I_{Ca} were made with the whole-cell configuration of the patch clamp technique using the solutions, apparatus and analysis programs previously described²⁹. All records were corrected for the current remaining in 100 μ M Cd²⁺. A prepulse protocol was used to determine facilitation in cells held at -80 mV. First, a 10-ms control pulse to 0 mV was applied (test 1) and then the cell was held at -80 mV for 2 s. Next a 25-ms conditioning prepulse to $+125$ mV was applied, then the cell was stepped back to -80 mV for 15 ms, and finally a second 10-ms test pulse (test 2) to 0 mV was applied. To emphasize changes in voltage-dependent activation of the currents, I_{Ca} was measured as the mean current from 5 to 6 ms after the beginning of the depolarization, which was before peak I_{Ca} was reached. The facilitation ratio was the ratio of I_{Ca} during test 2 divided by I_{Ca} during test 1.

release $\beta\gamma$ subunits, modulation of Ca^{2+} channels by these subunits provides a common mechanism through which many different neurotransmitters and G-protein types can cause modulation. As N-type and P/Q-type Ca^{2+} channels initiate transmitter release at most synapses, the flexible modulatory mechanism we have described is likely to have an important influence on the efficiency of synaptic transmission throughout the nervous system. □

Received 30 November 1995; accepted 15 February 1996.

1. Hille, B. *Trends Neurosci.* **17**, 531–536 (1994).
2. Hescheler, J. & Schultz, G. *Curr. Opin. Neurobiol.* **3**, 360–367 (1993).
3. Wickman, K. D. & Clapham, D. E. *Curr. Biol.* **5**, 278–285 (1995).
4. Dolphin, A. C. *Expl. Physiol.* **80**, 1–36 (1995).
5. Marchetti, C., Carbone, E. & Lux, H. D. *Pflügers Arch.* **406**, 104–111 (1986).
6. Bean, B. P. *Nature* **340**, 153–156 (1989).
7. Mintz, I. M. & Bean, B. P. *Neuron* **10**, 889–898 (1993).
8. Robitaille, R., Adler, E. M. & Charlton, M. P. *Neuron* **5**, 773–779 (1990).
9. Westenbroek, R. E. et al. *J. Neurosci.* **15**, 6403–6418 (1995).
10. Himing, L. D. et al. *Science* **239**, 57–61 (1988).
11. Wheeler, D. B., Randall, A. & Tsien, R. W. *Science* **264**, 107–111 (1994).
12. Luebke, J. I., Dunlap, K. & Turner, T. J. *Neuron* **11**, 895–902 (1993).

13. Takahashi, T. & Momiyama, A. *Nature* **366**, 156–158 (1993).
14. Hescheler, J., Rosenthal, W., Trautwein, W. & Schultz, G. *Nature* **325**, 445–446 (1987).
15. Wilk-Blaszczak, M. A., Singer, W. D., Gutowski, S., Sternweis, P. C. & Belardetti, F. *Neuron* **13**, 1215–1224 (1994).
16. Zhu, Y. & Ikeda, S. R. *Neuron* **13**, 657–669 (1994).
17. Diversé-Pierluissi, M., Goldsmith, P. K. & Dunlap, K. *Neuron* **14**, 191–200 (1995).
18. Stea, A. et al. *Proc. natn. Acad. Sci. U.S.A.* **91**, 10576–10580 (1994).
19. Ellis, S. B. et al. *Science* **241**, 1661–1664 (1988).
20. Clapham, D. E. & Neer, E. J. *Nature* **365**, 403–406 (1993).
21. Hepler, J. R. & Gilman, A. G. *Trends bio. Sci.* **17**, 383–387 (1992).
22. Breitwieser, G. E. & Szabo, G. *Nature* **317**, 538–540 (1985).
23. Pfaffinger, P. J., Martin, J. M., Hunter, D. D., Nathanson, N. M. & Hille, B. *Nature* **317**, 536–538 (1985).
24. Reuveny, E. et al. *Nature* **370**, 143–146 (1994).
25. Kofuji, P., Davidson, N. & Lester, H. A. *Proc. natn. Acad. Sci. U.S.A.* **92**, 6542–6546 (1995).
26. Wickman, K. D. et al. *Nature* **368**, 255–257 (1994).
27. Chen, J. et al. *Science* **268**, 1166–1169 (1995).
28. Hockerman, G. H., Johnson, B. D., Scheuer, T. & Catterall, W. A. *J. biol. Chem.* **270**, 22119–22122 (1995).
29. Bernheim, L., Beech, D. J. & Hille, B. *Neuron* **6**, 859–867 (1991).
30. Shapiro, M. S., Wollmuth, L. P. & Hille, B. *J. Neurosci.* **14**, 7109–7116 (1994).

ACKNOWLEDGEMENTS. We thank N. Nathanson and M. Simon for cDNA clones encoding the G-protein subunits, T. Snutch, K. Campbell, M. Harpold and A. Schwartz for cDNAs encoding the Ca^{2+} -channel subunits, D. Clapham for purified bovine brain $\beta\gamma$ protein, and S. Ikeda for advice on cell injection. This work was supported by the NIH, the W. M. Keck Foundation, DGAPA, the Alexander von Humboldt Foundation and Deutsche Forschungsgemeinschaft.

Inactivation of p27^{Kip1} by the viral E1A oncoprotein in TGF β -treated cells

Asoke Mal, Randy Y. C. Poon*, Philip H. Howe†, Hideo Toyoshima*, Tony Hunter* & Marian L. Harter‡

Departments of Molecular Biology and † Cell Biology, Cleveland Clinic Research Institute, 9500 Euclid Avenue, Cleveland, Ohio 44195, USA

* The Salk Institute, La Jolla, California 92037, USA

THE adenovirus oncoprotein E1A and the simian virus SV40 large T antigen can both reverse the strong growth-inhibitory effect of transforming growth factor(TGF)- β on mink lung epithelial cells^{1,2}: exposure of TGF- β causes these cells to arrest late in the G1 phase of the cell cycle (ref. 3). This arrest correlates with an increase in expression of the protein $\text{p15}^{\text{Ink4B}}$ (ref. 4), inactivation of the cyclin E/A-cdk2 complex by the inhibitory protein p27^{Kip1} (refs 5–7), and with the accumulation of unphosphorylated retinoblastoma protein⁸. The rescue by E1A of cells from TGF- β arrest is partly independent of its binding to retinoblastoma protein¹. Here we show that E1A directly affects the cyclin-dependent kinase inhibitor p27^{Kip1} in TGF- β -treated cells by binding to it and blocking its inhibitory effect, thereby restoring the activity of the cyclin-cdk2 kinase complex. In this way, E1A can overcome the effect of TGF- β and modulate the cell cycle. To our knowledge, E1A provides the first example of a viral oncoprotein that can disable a cellular protein whose function is to inhibit the activity of cyclin-dependent kinases.

As microinjection of a purified, bacterially produced E1A protein⁹ can rescue mink lung epithelial (Mv1Lu) cells from growth inhibition by TGF- β (results not shown), and as TGF- β -treated Mv1Lu cells show decreased cyclin E/A-cdk2 activity^{5–7}, we investigated whether E1A could restore activity to these complexes. As expected, cdk2 immune complexes from extracts of proliferating cells had associated histone H1 kinase activity, whereas quiescent or TGF- β -treated cells had very little associated activity (Fig. 1a)^{5,6,10}. However, addition of E1A to extracts of TGF- β -treated cells restored cyclin-cdk2 kinase activity in cdk2 immune complexes to a level comparable to that of cyclin-cdk2 complexes from untreated cells (Fig. 1b). When incremental

amounts of E1A were added to fixed amounts of TGF- β -treated cell extract, cdk2 H1 kinase activity increased linearly with a dose (Fig. 1c), suggesting that E1A is titrating out an inhibitory activity associated with cyclin-cdk2 complexes. We conclude that E1A can reverse a TGF- β -induced decrease in cyclin-cdk2 activity *in vitro*.

Because cyclin E-cdk2 complexes in TGF- β -treated Mv1Lu cells are inactivated by binding to p27^{Kip1} protein *in vitro*⁶, we speculated that E1A might be abrogating the inhibitory effect of p27^{Kip1} . We therefore devised an *in vitro* assay based on the fact that p27^{Kip1} is heat-stable^{6,10,11} whereas E1A is not⁹. E1A was incubated in extracts of TGF- β -treated cells and recovered by immunoprecipitation with M73 monoclonal antibody¹². Precipitated E1A, together with any associated proteins, was released from protein A-Sepharose beads and mixed with extracts of proliferating cells. This mixture was then immunoprecipitated with anti-cdk2 and the resulting immune complexes assayed for H1 kinase activity. The released products contained an inhibitor of cyclin-cdk2 kinase (Fig. 2, lane 3), as judged by comparison with the activity of cyclin-cdk2 isolated from untreated proliferating cell extracts (lane 1). Furthermore, when we mixed a TGF- β -treated cell extract, to which E1A had been added and then removed, with equal amounts of proliferating-cell extract, the cyclin-cdk2 immune complexes showed little decrease in H1 kinase activity (lane 4). We confirmed that the TGF- β -treated extracts contained an inhibitory activity before exposure to E1A by mixing this extract with proliferating-cell extract (1:1) and found that the kinase activity of cyclin-cdk2 immunoprecipitates was dramatically reduced (lane 2). As expected, E1A alone did not inhibit the activity of cyclin-cdk2 complexes in extracts of proliferating cells (data not shown).

To test whether the inhibitor of cyclin-cdk2 catalytic activity that is removed from TGF- β -treated Mv1Lu cells by E1A was p27^{Kip1} , we added E1A to extracts of TGF- β -treated cells to see whether E1A could bind p27^{Kip1} (Fig. 3a): p27^{Kip1} was readily detectable in E1A immunoprecipitates by blotting with anti- p27 antibodies. We also found that when E1A was incubated with a glutathione-S-transferase (GST)- p27^{Kip1} fusion protein⁷ (Fig. 3b), it bound specifically to p27^{Kip1} but not to GST or GST-cdk2. We next investigated the region on p27^{Kip1} for binding of E1A and found that a p27 mutant containing N-terminal residues 1–91 bound to E1A much less efficiently than a p27 mutant containing C-terminal residues 91–197 (Fig. 3b). We conclude that E1A interacts with p27^{Kip1} through sequences residing mainly in the C-terminal half of p27^{Kip1} .

To test whether E1A can directly reverse the inhibitory effect of p27^{Kip1} on cyclin E/A-cdk2, we used a purified system (Fig. 3c).

‡ To whom correspondence should be addressed.

The activity of a GST-cdk2-cyclin A complex was inhibited by p27 but was almost fully restored after addition of E1A; E1A had little effect on the activity of GST-cdk2-cyclin A alone. We found that E1A displaced (Fig. 3c) about half of the p27 associated with the GST-cdk2-cyclin A complex. Under these conditions, E1A showed no intrinsic binding specificity for the complex in either the p27-bound or unbound state (data not shown).

Our *in vitro* results suggest that E1A, in part, overcomes TGF- β suppression of cell growth by binding to and disabling p27^{Kip1}. Can E1A interfere with the activities of p27^{Kip1} *in vivo*? Transfection of

TGF- β -treated cells with a plasmid containing E1A complementary DNA¹³ restored histone H1 kinase activity to cyclin-cdk2 complexes (Fig. 4a, lane 2), whereas a control plasmid did not (Fig. 4a, lane 1). A phosphorylated form of the retinoblastoma protein (Rb) was also observed in E1A-transfected cells (Fig. 4a, lane 4), but not in control cells (Fig. 4a, lane 3). We next determined whether E1A was reducing the amount of p27^{Kip1} associated with cyclin-cdk2 complexes *in vivo* (Fig. 4b) and found that the level of p27^{Kip1} in anti-cdk2 immunoprecipitates from cells transfected with E1A much less than in control cells

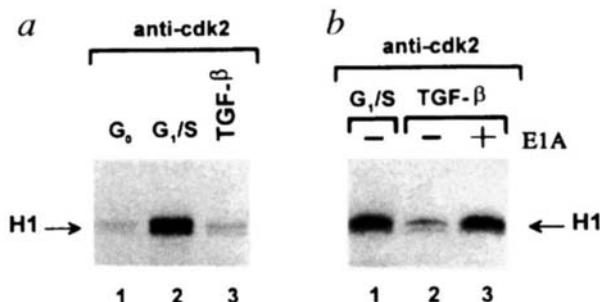
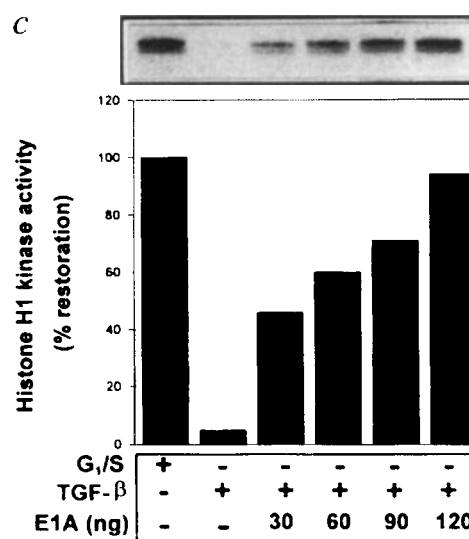


FIG. 1 E1A restores cdk2-associated kinase activity in TGF- β -treated cells. **a**, Quiescent Mv1Lu cells were restimulated with serum for 15 h (G1/S) either in the presence or absence of TGF- β . Whole-cell extracts were prepared and immunoprecipitated with an anti-cdk2 antibody. Cdk2-associated kinase activity in the immunoprecipitates was determined using histone H1 as substrate. **b**, As **a**, except that the extracts, with one from TGF- β -treated cells containing added E1A (200 ng), were incubated on ice for 30 min before immunoprecipitation and the kinase reaction. **c**, Extracts of TGF- β -treated cells with or without the indicated amounts of recombinant E1A were incubated in parallel along with proliferating cells. Cdk2 was then recovered by immunoprecipitation and assayed for associated kinase H1 activity. Radiolabelled histone H1 viewed on SDS-polyacrylamide gels was quantified (in relative units) with a phosphorimager and ImageQuant software (Molecular Dynamics) to calculate the percentage of restoration. **METHODS**. An authentic E1A 243-amino-acid protein produced and purified from bacteria is reported elsewhere⁹. Cells were grown to confluency in DMEM medium containing 10% FBS and thereafter in medium without serum to achieve quiescence (G0)^{19,20}. Cells were then restimulated with serum (10% FBS) and EGF (20 ng ml⁻¹) for 15 h (G1/S)³ either in the presence or absence of TGF- β (5 ng ml⁻¹). Whole-cell extracts were



prepared accordingly²¹. Extracts (100 μ g) of quiescent, proliferating, or TGF- β -treated cells were immunoprecipitated^{22,23} with anti-cdk2 (Santa Cruz Biotechnology). For assay of kinase activity, immune complexes of cyclin-cdk2 were resuspended in 15 μ l buffer²² containing histone H1 (0.2 μ g), 50 μ M cold ATP, and 5 μ M [γ -³²P]ATP. After incubating for 30 min at 30 $^{\circ}$ C, the reactions were terminated by 2 $\times$ sample buffer. Phosphorylated products were analysed on an 8% SDS-polyacrylamide gel and visualized by autoradiography. Exposure time in all experiments was about 1 h. Under no circumstances was purified E1A able to phosphorylate histone H1 or any other substrate in a kinase reaction (data not shown).

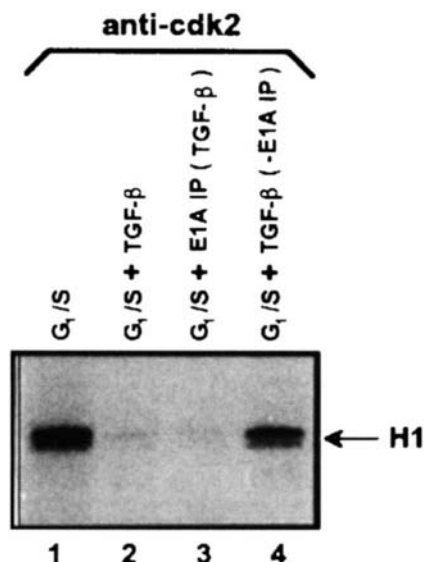


FIG. 2 E1A removes an inhibitory activity from TGF- β -treated cells. Quiescent Mv1Lu cells were restimulated as described for Fig. 1 and grown for 15 h either with or without TGF- β . As a control, cdk2-associated histone H1 kinase activity was determined in 100 μ g G1/S extract (lane 1) after immunoprecipitation with antibodies against cdk2. Lane 2, same assay, except that cdk2-associated histone H1 kinase activity was determined after mixing TGF- β extracts (100 μ g) with G1/S extracts (100 μ g). Lane 3, cdk2-associated histone H1 kinase activity was measured in G1/S extract after mixing with proteins released from E1A immune complexes¹⁰ recovered from TGF- β extracts after addition of E1A. Lane 4, cdk2-associated histone H1 activity in mixed G1/S and TGF- β extracts; the latter had formerly contained E1A.

METHODS. Culture and treatment of Mv1Lu cells with or without TGF- β , and preparation of whole-cell extracts are described for Fig. 1, except the extract in buffer²¹ contained NaCl at 150 mM instead of 250 mM. For the inhibitory reaction, E1A (200 ng) was incubated for 15 min at 30 $^{\circ}$ C in extracts of TGF- β -treated cells (100 μ g) and recovered by immunoprecipitation with an antibody specific for E1A (M73)¹². Protein A-Sepharose beads were treated accordingly¹⁰, quenched on ice, diluted in buffer²¹ (60-fold), and centrifuged for 1 min. Proteins released from beads were mixed with extracts (100 μ g) of proliferating cells and afterwards this mixture was assayed for cyclin-cdk2-associated histone H1 kinase activity after immunoprecipitation (Fig. 1). To assay for any inhibitory activity, SDS was added to extracts of proliferating cells to a final concentration of 0.033%, an amount of SDS that did not affect the catalytic activity of cyclin-cdk2 (unpublished results).

(lane 2 versus 3), consistent with the *in vitro* results (Fig. 3c). p27^{Kip1} was present in E1A immune complexes from TGF- β -treated cells transfected with the E1A plasmid, but not the control plasmid (Fig. 4c: compare lanes 1 and 2). Transient transfection of E1A into TGF- β -treated cells reactivates cyclin-cdk2 kinase activity and this may be responsible for the phosphorylation of Rb. Moreover, E1A can also associate with p27^{Kip1} and reduce the level of p27^{Kip1} associated with cyclin-cdk2 complexes *in vivo*.

This ability of E1A to displace p27^{Kip1} from cyclin-cdk2 complexes and restore their catalytic activity is one mechanism by

which E1A may overcome the growth-inhibitory effects of TGF- β in mink lung cells. This may complement the ability of E1A to disrupt Rb-E2F¹⁴ and cyclin-cdk2-E2F-p107/130 complexes¹⁵ and release E2F, a transcription factor important in the activation of S-phase genes¹⁶.

As TGF- β -induced cell-cycle arrest results, in part, from inactivation of cyclin E/A-cdk2 complexes due to increased binding of p27^{Kip1}^{4,10}, and as E1A does not need to bind to Rb to counteract the effects of TGF- β ¹, the ability of E1A to reverse p27^{Kip1} inhibition of cyclin-cdk2 activity may be critical in allowing

FIG. 3 E1A binds to p27^{Kip1} and restores kinase activity to GST-cdk2-cyclin A-p27 in a purified system. **a**, Quiescent Mv1Lu cells were restimulated for 15 h and in the presence of TGF- β (5 ng ml⁻¹). Extracts (100 μ g) were incubated with or without E1A (200 ng) and then immunoprecipitated with M73 or anti-p27^{Kip1} (ref. 18), respectively. Precipitates were electrophoresed alongside 60 μ g of straight extract, transferred to a PVDF membrane, and detected with anti-p27^{Kip1} and ECL (Amersham). **b**, GST, GST-cdk2, GST-p27, GST-p27C or GST-p27N, immobilized on glutathione beads, were incubated with purified E1A (75 ng). Input refers to 100% of the protein used in the binding assay. **c**, Purified GST-cdk2 and His-tagged cyclin A were immobilized onto GSH-agarose beads. One set of complexes were incubated in buffer with purified His-tagged p27 (lanes 2 and 3) and another in buffer alone (lanes 1 and 4). Washed complexes were then incubated with (lanes 3 and 4) or without purified E1A (lanes 1 and 2). Immobilized complexes, in duplicate, were assayed for histone H1 kinase activity and for the presence of p27 by immunoblotting. **METHODS.** Construction and purification of GST-cdk2, GST-p27, GST-p27C, GST-p27N, His-tagged p27 and His-tagged cyclin A (consisting of the regulatory subunit) have been described^{6,7,24}, as have the binding of equimolar amounts of GST fusion proteins to glutathione beads, incubation with E1A, and E1A detection by immunoblotting after elution¹³. To detect the restoration of activity in a purified system, GST-cdk2 and His-tagged cyclin A (500 ng each) were initially activated by CAK-immunoprecipitates^{25,26} and bound to GSH-agarose beads. Beads were preincubated²⁵ with or without His-tagged p27 (2 μ g), washed, and incubated with or without purified E1A (1 μ g). Washed beads were either assayed for histone H1 kinase²⁶ or suspended in SDS-PAGE loading buffer for elution of bound products. Eluted products were immunoblotted for p27 or E1A as described. Excess p27 and E1A were added to GST-cdk2-cyclin A to maximize the inhibition or restoration of kinase activity, respectively.

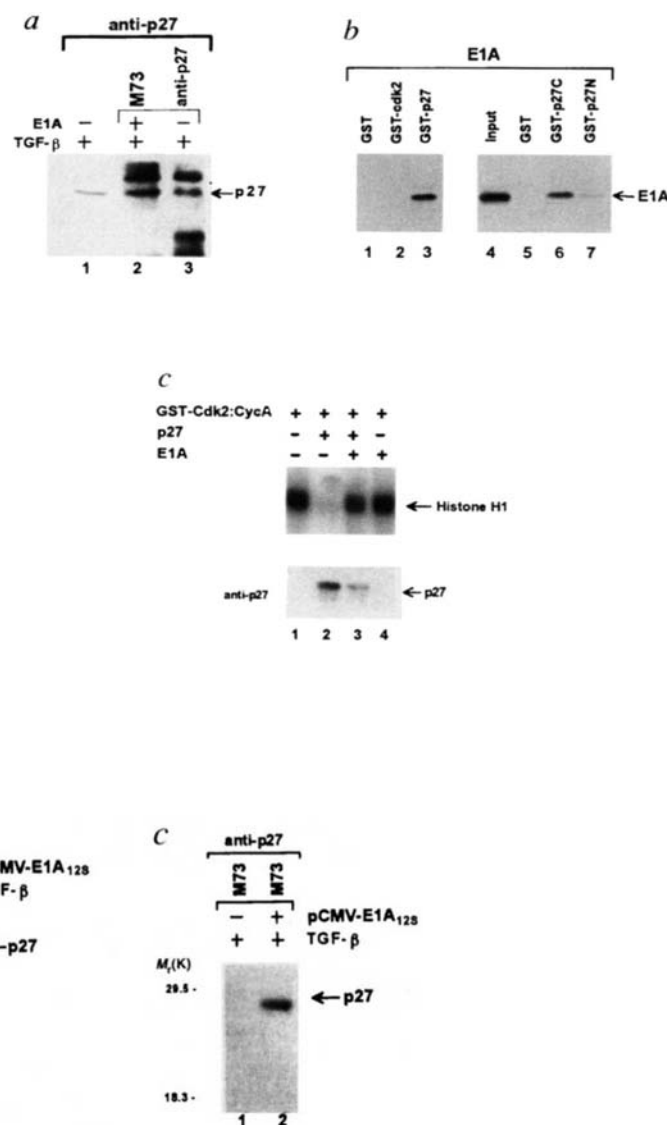


FIG. 4 E1A associates with p27^{Kip1} and restores kinase activity to cyclin-cdk2 complexes *in vivo*. Mv1Lu cells were transfected with pCMV-E1A_{12S} (where 12S refers to the size of E1A) or control plasmid (pCDNA3) and treated with TGF- β (10 ng ml⁻¹) and EGF (20 ng ml⁻¹) in fresh medium containing 10% FBS. After 24 h, cells were collected and extracts prepared. Equal fractions of each transfection (representing protein recovered from the same number of cells) were analysed in all experiments. **a**, Cdk2-associated kinase histone H1 activity in transfected TGF- β -treated cells was determined as for Fig. 1. For analysis of Rb, extracts (100 μ g) of transfected cells were immunoprecipitated with anti-Rb antibody (C-15; Santa Cruz Biotechnology) and precipitates analysed on a 7.5% SDS-polyacrylamide gel followed by immunoblotting and probing with an Rb-specific monoclonal antibody (IF8; Santa Cruz) and ECL detection. **b**, pCMV-E1A or pCDNA3 whole-cell extracts (450 μ g) were immunoprecipitated with antibodies against cdk2 or non-immune rabbit serum (NRS). The precipitated samples were resolved on a 10% SDS-polyacrylamide gel. Separated products were

transferred to a PVDF membrane and probed with anti-p27^{Kip1}, followed by ECL. **c**, Whole-cell extracts of pCMV-E1A or pCDNA3 were immunoprecipitated with M73 and the precipitated products resolved on a 12% SDS-polyacrylamide gel. Bound p27^{Kip1} was visualized by immunoblot analysis using antibodies against p27^{Kip1}. Molecular mass markers are shown on the left.

METHODS. A cDNA fragment containing the full E1A 12S coding sequence was ligated into pCDNA3 plasmid (Invitrogen), downstream of the cytomegalovirus promoter to construct pCMV-E1A (ref. 13). Mv1Lu cells at 70% confluency were transfected with 15 μ g pCMV-E1A or pCDNA3 by a standard calcium phosphate precipitation method (GIBCO/BRL). After removal of DNA precipitates, cells were treated with 10 ng ml⁻¹ TGF- β and 20 ng ml⁻¹ EGF in fresh medium containing 10% FBS. After 24 h, cells were collected for preparation of whole-cell extracts as described for Fig. 1. Transfection efficiency under our conditions was ~40%.

TGF- β -arrested cells to enter and progress through S phase, a process requiring cyclin-cdk2 activity¹⁷. The fact that binding of E1A to the C-terminal domain of p27 is stronger than to the N-terminal inhibitory domain indicates that E1A may bind to the free C-terminal part of p27 in a cyclin-cdk2-p27 complex and is in a position competitively to displace the N-terminal domain from cyclin-cdk2. Once displaced (Fig. 3c, 4b), p27^{Kip1} could be sequestered by E1A, preventing it from rebinding and abrogating its inhibitory effect. To our knowledge, this binding of E1A to block the action of p27^{Kip1} provides the first example of a viral oncoprotein targeting and activating a cellular protein whose function is negatively to regulate cyclin-dependent kinases. □

Received 8 December 1995; accepted 25 January 1996.

1. Abraham, S. E., Carter, M. C. & Moran, E. *Molec. Biol. Cell* **3**, 655–665 (1992).
2. Pietsch, J. A. *et al. Cell* **61**, 777–785 (1990).
3. Howe, P. H., Draetta, G. & Leof, E. B. *Molec. cell. Biol.* **11**, 1185–1194 (1991).
4. Reynisdottir, I., Polyak, K., Lavarone, A. & Massague, J. *Genes Dev.* **9**, 1831–1845 (1995).
5. Koff, A., Ohtsuki, M., Polyak, K., Roberts, J. M. & Massague, J. *Science* **260**, 536–539 (1993).
6. Polyak, K. *et al. Genes Dev.* **8**, 9–22 (1994).
7. Toyoshima, H. & Hunter, T. *Cell* **78**, 67–74 (1994).

8. Laiho, J., DeCaprio, J., Ludlow, J., Livingston, D. & Massague, J. *Cell* **62**, 175–185 (1990).
9. Wang, D.-M., Dali, B. & Harter, M. L. *Protein Express. Purif.* **3**, 8–17 (1992).
10. Slingerland, J. M. *et al. Molec. cell. Biol.* **14**, 3683–3694 (1994).
11. Polyak, K. *et al. Cell* **78**, 59–66 (1994).
12. Harlow, E., Franza, B. R. Jr & Schley, C. J. *Virology* **55**, 533–546 (1985).
13. Lundblad, J. R., Kwok, R. P. S., Lurance, M. E., Harter, M. L. & Goodman, R. H. *Nature* **374**, 85–88 (1995).
14. Nevins, J. R. *Biochem. Soc. Trans.* **21**, 935–938 (1993).
15. Hunter, T. & Pines, J. *Cell* **79**, 573–582 (1994).
16. Sherr, C. J. *Cell* **79**, 551–555 (1994).
17. Hunter, R. *Cell* **75**, 839–841 (1993).
18. Kato, J.-Y., Matsuo, M., Polyak, K., Massague, J. & Sherr, C. J. *Cell* **79**, 487–496 (1994).
19. Ewen, M. E., Sluss, H. K., Whitehouse, L. L. & Livingston, D. M. *Cell* **74**, 1009–1020 (1993).
20. Howe, P. H., Dobrowolski, S. F., Reddy, K. B. & Stacey, D. W. *J. Biol. Chem.* **268**, 21448–21452 (1993).
21. Stacey, D. W., Dobrowolski, S. F., Piotrowski, A. & Harter, M. L. *EMBO J.* **13**, 6107–6114 (1994).
22. Banerjee, A. C. *et al. Oncogene* **9**, 1733–1737 (1994).
23. Faha, B., Harlow, E. & Lees, E. *J. Virol.* **67**, 2456–2465 (1993).
24. Poon, R. Y. C. & Hunter, T. *Science* **270**, 90–93 (1995).
25. Poon, R. Y. C., Yamashita, K., Adamczewski, J. P., Hunt, T. & Shuttleworth, J. *EMBO J.* **12**, 3123–3132 (1993).
26. Poon, R. Y. C. *et al. J. Cell Sci.* **107**, 2789–2799 (1994).

ACKNOWLEDGEMENTS. We thank J. Massague for anti-p27^{Kip1} (ref. 18), C. Weyman for critical discussion and for comments on the manuscript, and D. Herzberg for editorial assistance in the preparation of this manuscript. This work was supported by grants from the NIH and the Cleveland Clinic Foundation.

Sequence-specific DNA binding by Ku autoantigen and its effects on transcription

Ward Giffin*, Heather Torrance†, David J. Rodda†, Gratien G. Préfontaine†, Louise Pope* & Robert J. G. Haché*†‡

Departments of * Medicine and † Biochemistry, University of Ottawa, Loeb Institute for Medical Research, Ottawa Civic Hospital, 1053 Carling Avenue, Ottawa, Ontario K1Y 4E9, Canada

DNA-DEPENDENT protein kinase (DNA-PK) has been implicated in several nuclear processes including transcription^{1–3}, DNA replication^{4,5}, double-stranded DNA break repair, and V(D)J recombination^{6–10}. Linkage of kinase and substrate on DNA *in cis* is required for efficient phosphorylation¹¹. Recruitment of DNA-PK to DNA is by Ku autoantigen, a DNA-end-binding protein required for DNA-PK catalytic activity¹¹. Although Ku is known to translocate along naked DNA¹², how DNA-end binding by Ku might lead to DNA-PK-mediated phosphorylation of sequence-specific DNA-binding proteins *in vivo* has not been obvious. Here we report the identification of Ku as a transcription factor that recruits DNA-PK directly to specific DNA sequences. NRE1 (negative regulatory element 1) is a DNA sequence element (–394/–381) in the long terminal repeat of mouse mammary tumour virus (MMTV) that is important for repressing inappropriate viral expression^{13–16}. We show that direct binding of Ku/DNA-PK to NRE1 represses glucocorticoid-induced MMTV transcription.

Fractionation of a nuclear extract from Jurkat T cells to identify NRE1-binding factors resulted in two peptides of relative molecular mass 70,000 and 85,000 (*M_r* 70K and 85K, respectively) being purified to near homogeneity (Fig. 1a). Microsequence analysis of the amino terminus of the 85K peptide revealed identity to Ku¹⁷ (p70/p86). Monoclonal antibodies confirmed the NRE1 binding activity in the purified fraction and in crude nuclear extracts as Ku (Fig. 1b).

The apparent specificity of Ku for NRE1 was striking. DNase I

footprinting (Fig. 1c) showed that a region of the MMTV long terminal repeat (LTR) centred over NRE1 remained completely protected from DNase I digestion following the competition of DNA-end binding with highly sheared calf thymus DNA (Fig. 1c, lanes 3 and 7). However, it remained possible that protection of NRE1 by Ku resulted from the stable accumulation of translocated Ku following an initial interaction with the DNA ends.

Direct binding of purified and recombinant Ku to NRE1 was demonstrated using covalently closed DNA microcircles and plasmid DNAs (Fig. 2). Microcircles and plasmids were predigested with exonuclease III, S1 nuclease and *Bal31* to exclude the possibility of Ku binding to nicks, ends or structural features in the DNA¹⁸ (Fig. 2a, and data not shown). A Ku–DNA complex was observed on NRE1-containing microcircles (Fig. 2b, lanes 1–7), but not control circles (lanes 8–11); see lanes 2 and 6. Complex formation was competed by closed circular pBluescript containing a single copy of NRE1 (pNRE1) and NRE1 oligonucleotide, but not by the pBluescript parent plasmid or DNA ends (Fig. 2b, lanes 2–5). In footprinting experiments (Fig. 2c), protection of LTR sequences over NRE1 (lanes 3 and 8) corresponded closely to that observed above on linear DNA (Fig. 1c, lanes 3, 4, 7 and 8). Binding was competed by a 20-fold excess of closed circular pNRE1 (lanes 5 and 10), but not by pBluescript (lanes 4 and 9). We concluded that NRE1 is a direct, internal, sequence-specific binding site for Ku antigen that is strongly preferred to DNA ends.

Importantly, NRE1 was not a static Ku binding site. Addition of Mg²⁺ to incubations of Ku with microcircles promoted the translocation of Ku from NRE1 (Fig. 1c and Fig. 2d). In electrophoretic mobility-shift assay (EMSA), translocation of Ku from NRE1 resulted in several typical¹² lower-mobility Ku/DNA complexes (Fig. 2d, lane 4). Similarly, in footprinting experiments on linear DNA, early addition of Mg²⁺ resulted in the appearance of hypersensitive sites representative of translocating Ku¹¹ in the absence of detectable DNA-end binding (Fig. 1c, lanes 5 and 9).

Although DNA-PK phosphorylates many sequence-specific DNA-binding proteins *in vitro*^{2,4,5}, its ability to phosphorylate the same factors *in vivo* and affect their regulatory potential remains hypothetical. Glucocorticoid receptor and octamer transcription factor 1 (Oct-1) are two potential DNA-PK targets that regulate MMTV transcription^{19,20} (Fig. 3a). Oct-1 and a glutathione *S*-transferase (GST)–Oct-1 fusion protein are phosphorylated by DNA-PK *in vitro*⁴, and potential DNA-PK phosphorylation sites⁴ also occur on the glucocorticoid receptor.

In striking contrast to previous results obtained with covalently closed circular DNAs¹¹, purified DNA-PK²¹ was fully active on

‡ To whom correspondence should be addressed.

relaxed circular and supercoiled MMTV LTR DNA (Fig. 3b, c), phosphorylating both recombinant glucocorticoid receptor and GST-Oct-1 (lanes 6 and 8). Deletion of NRE1, however, resulted in a plasmid that only supported DNA-PK phosphorylation when linear (lanes 5, 7 and 9). Phosphorylation was dependent upon DNA binding by glucocorticoid receptor and GST-Oct-1, as DNA-PK was unable to phosphorylate either factor in the absence of glucocorticoid receptor and Oct-1 binding sites, and a glucocorticoid receptor point mutant unable to bind a glucocorticoid regulatory element, C495Y²², was not phosphorylated using any DNA (data not shown).

These experiments also demonstrated that at least one previously proposed sequence-specific Ku binding site, the octamer motif²³, of which two copies occur in pHC364, was unable to direct DNA-PK activity (Fig. 3b, c, lanes 7 and 9). We also failed to detect binding of Ku to octamer motif-containing microcircles (data not shown). Further, the lack of obvious sequence homology between NRE1 and the other proposed Ku binding sites suggests that most will prove to be pause sites for translocating Ku, rather than direct binding sites.

The participation of Ku/DNA-PK in NRE1-mediated transcriptional effects was evaluated by transient transfection assays in fibroblasts containing inactivating mutations in the p86 subunit of Ku or the catalytic subunit of the kinase (Fig. 4). In parental V79 cells (Fig. 4a), NRE1 repressed the activation of MMTV transcription by glucocorticoids almost 10-fold (lanes 1 and 2). By

contrast, in clonally derived p86⁻ XR-V15B cells lacking detectable Ku⁶, NRE1 very slightly activated transcription (lanes 3 and 5). Introduction of exogenous p86 by co-transfection of a p86 cDNA expression plasmid⁶ rescued NRE1-mediated repression, but had no effect on MMTV expression in the absence of NRE1 (lanes 4 and 6). Thus Ku was required for the repression of MMTV transcription through NRE1.

Transfections were repeated on severe combined immunodeficient (SCID) mouse fibroblast cells containing a mutation that eliminates DNA-PK activity without affecting Ku status^{9,10,24} (Fig. 4b). In the wild-type CB-17 cells, MMTV transcription was repressed almost fourfold by NRE1 (lanes 1 and 2). By contrast, no repression was detected in the DNA-PK-deficient Sf-7 clone (lanes 3 and 4). Thus it appears that Ku is not sufficient for NRE1 regulatory effects, but requires the active participation of DNA-PK.

Our results demonstrate that NRE1 is a high-affinity site for the recruitment of Ku/DNA-PK to DNA. We believe that such sequence-specific targeting of a kinase is without precedent, and has important implications for the nuclear processes in which Ku/DNA-PK have been implicated. In particular, it suggests a novel mechanism for promoter/enhancer specific modification of transcription factor activity. The phosphorylation of glucocorticoid receptor and Oct-1 on the MMTV promoter by DNA-PK *in vitro* suggests two obvious candidates for transmission of the regulatory signal on the MMTV LTR. Intriguingly, restricting DNA-PK

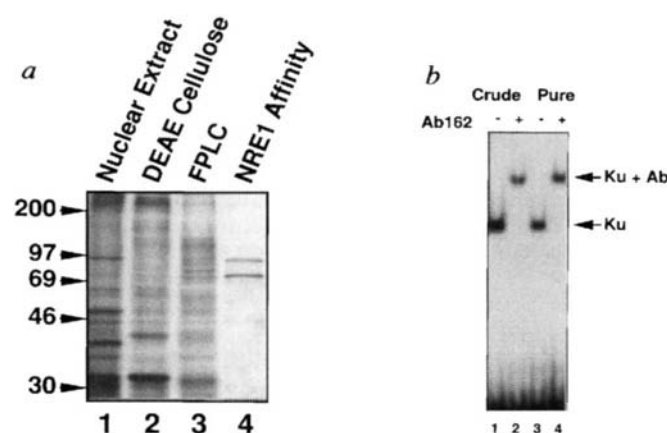
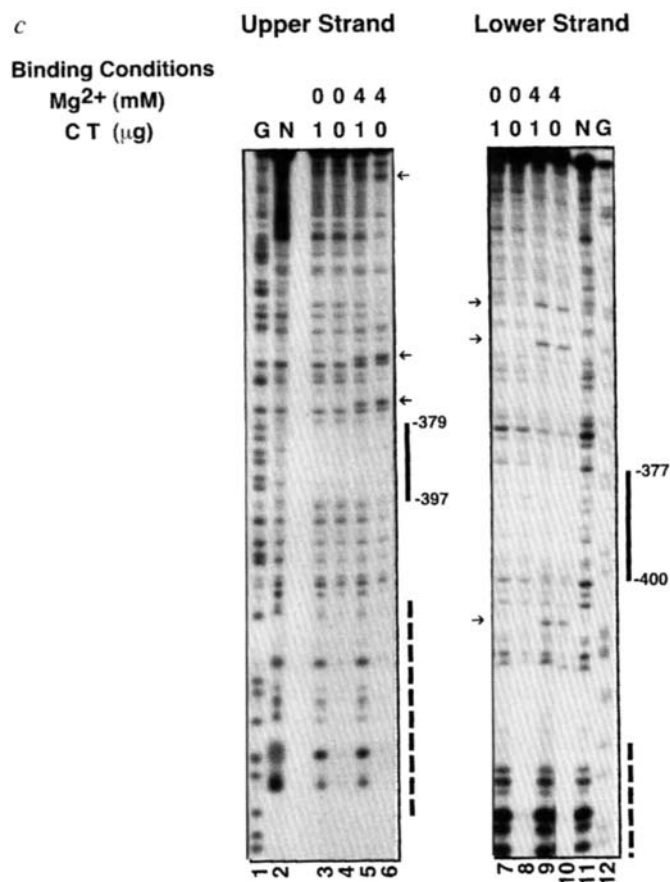


FIG. 1 Identification of Ku antigen as an NRE1 binding protein. **a**, Silver-stained 10% SDS polyacrylamide gel of column fractions containing NRE1 binding activity. Lane 1, Jurkat nuclear extract; lane 2, 0.2 M KCl DEAE cellulose fraction; lane 3, Superose 12 fast protein liquid chromatography (FPLC) fraction 6; lane 4, 0.5 M KCl NRE1-Sepharose affinity fraction. **b**, EMSA with Jurkat nuclear extract (lanes 1 and 2) and purified Ku fraction from **a** (lanes 3 and 4). In lanes 2 and 4, Ku monoclonal antibody (Ab) 162 (ref. 25) was added to the incubation. A second Ku monoclonal antibody, N3H10 (ref. 25), but not nonspecific antibodies, supershifted the same complex (data not shown). **c**, DNase I footprinting and the effect of Mg²⁺ on purified Ku binding to a -421/-106 MMTV DNA fragment. Binding was performed in the presence (lanes 3, 5, 7 and 9) or absence (lanes 4, 6, 8, 10) of 1 μg sheared calf thymus DNA (CT, average size 400 base pairs). MgCl₂ (4 mM) was added at the time of DNase I addition (lanes 3, 4, 7 and 8) or at the beginning of the incubation (lanes 5, 6, 9 and 10). Lanes 1 and 12, G reference; lanes 2 and 11, naked (N) DNA digested with DNase I. The solid bars mark protection over NRE1, the dashed bars mark protection of DNA ends. The arrows indicate DNase I hypersensitive sites typical of translocating Ku.

METHODS. Jurkat T-cell nuclear extract¹³ was passed over DEAE-cellulose, Superose 12 FPLC (Pharmacia), and NRE1-Sepharose. A 26-bp NRE1-containing oligonucleotide (upper strand, 5'-ACCGGACTGAGAAAGAGAAAGACGAC-3', 5 bp overhang), ligated to an average repeat length of 10, was used to prepare the affinity column. Before affinity chromatography, 300 μg sheared calf thymus DNA was added to the sample (final volume 12 ml).



Columns were equilibrated with (in mM) 20 HEPES, pH 7.9, 20% glycerol, 100 KCl, 0.2 EDTA, 0.2 phenylmethylsulphonyl fluoride, 0.5 DTT, 0.1% Noridet-PHO. NRE1 binding was followed by EMSA¹³. The N-terminal sequence of the 85K protein was VRSGNKAAVV, identical to amino acids 2-11 of p86 Ku, 1-MVRSNGKAVV-11 (ref. 17). DNase I footprinting was performed as previously described¹³.

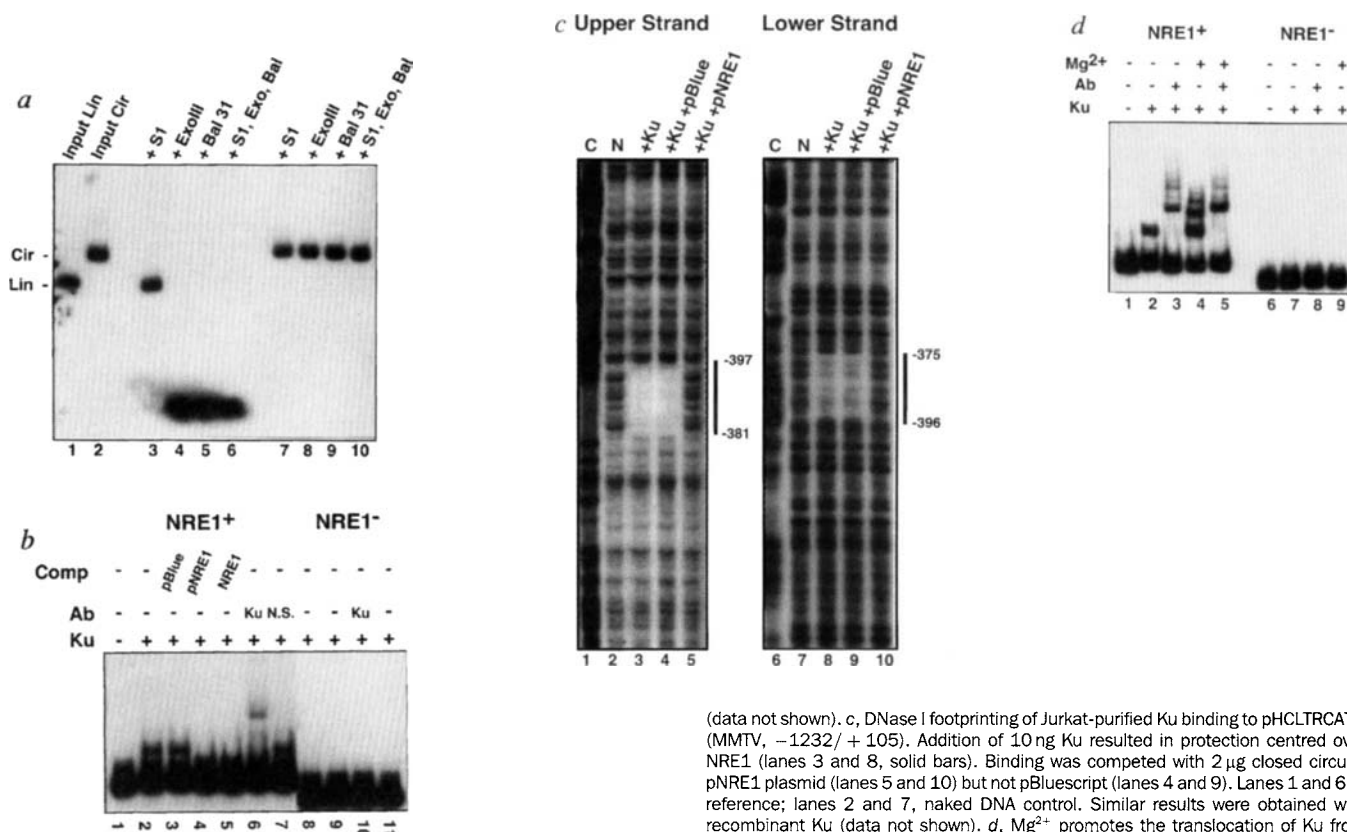
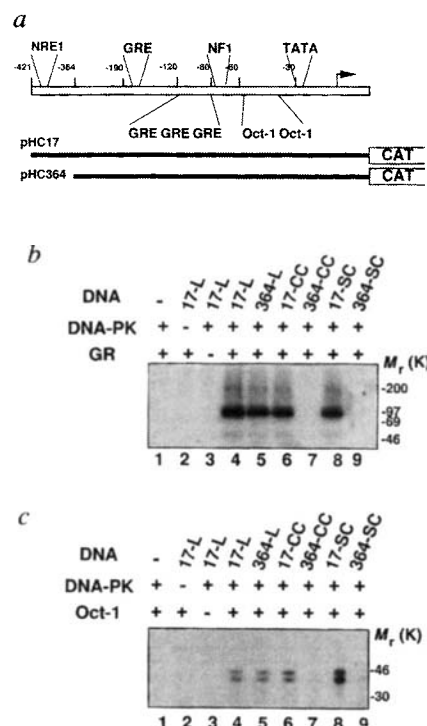


FIG. 2 Ku binds specifically to and translocates from NRE1. **a**, Polyacrylamide gel (4%) of linear (Lin, lanes 1 and 3–6) and covalently closed circular (Cir, lanes 2 and 7–10) 223-bp NRE1-containing fragments incubated with nucleases S1, ExoIII and Bal31, singly or in combination. Microcircles were resistant to all three nucleases. **b**, EMSA of recombinant Ku²⁶ binding to NRE1-containing microcircles. Lanes 1–7, 223-bp NRE1-containing microcircle; lanes 8–11, 200-bp control circle lacking NRE1. Addition of Ku (~1 ng), competitor DNAs (lane 3, 2 µg pBluescript; lane 4, 2 µg pNRE1; lane 5, 200 ng NRE1 oligonucleotide), and 100 ng anti-Ku antibody J15 (Ku, Santa Cruz) or 500 ng non-specific antibody BuGR (N.S., Affinity BioReagents) are indicated. Incubations included 200 ng sheared calf thymus DNA (100–200-fold excess of DNA ends). No binding was observed with extract from mock-infected cells

(data not shown). **c**, DNase I footprinting of Jurkat-purified Ku binding to pHLTRCAT¹⁴ (MMTV, –1232/+105). Addition of 10 ng Ku resulted in protection centred over NRE1 (lanes 3 and 8, solid bars). Binding was competed with 2 µg closed circular pNRE1 plasmid (lanes 5 and 10) but not pBluescript (lanes 4 and 9). Lanes 1 and 6, C reference; lanes 2 and 7, naked DNA control. Similar results were obtained with recombinant Ku (data not shown). **d**, Mg²⁺ promotes the translocation of Ku from NRE1. EMSA performed as in **b**, with 1 ng of Jurkat-purified Ku, except that 4 mM MgCl₂ was added to the incubation in lanes 4, 5 and 9. METHODS. pNRE1 is pBlue (pBluescript, Stratagene) with a 23-bp NRE1-containing oligonucleotide¹³ in the SmaI site. Microcircles were prepared by ligation of PvuII/XhoI pNRE1 and pBlue fragments following fill in of the ends in the presence of ³²P dATP. Gel-purified microcircles were resistant to exonuclease III, Bal31 and S1 nuclease. Recombinant Ku was expressed from baculovirus and purified by nickel affinity chromatography²⁶. DNase I footprinting was performed as previously described¹³, except that 10 ng nuclease-resistant relaxed circular pHLTR plasmid¹⁴ was used in the presence of 200 ng sheared calf thymus DNA. Cleavage was visualized by primer extension of oligonucleotides specific for pHLTR (upstream, –464 5'-AACCTTTCCCAAAAGTGCATCC-3'; downstream, –313 3'-TGTATACAGATCCCTCCCTTCG-3').

FIG. 3 DNA-PK phosphorylates recombinant glucocorticoid receptor and Oct-1-POU domain fusion protein *in vitro*. **a**, Schematic representation of MMTV transcriptional regulatory elements present in plasmids pHC17 and pHC364 (ref. 14). Only pHC17 contains NRE1, but both contain octamer motifs. The MMTV transcriptional initiation site is indicated by the arrow. NRE1, sequence specific Ku binding site; GRE, glucocorticoid response element; Oct-1, octamer motif; NF1, nuclear factor 1 binding site; TATA, TATA box; CAT, chloramphenicol acetyltransferase reporter gene. **b**, Phosphorylation of human glucocorticoid receptor by DNA-PK. The components added to each incubation are summarized above the lanes. DNAs: 17, pHC 17; 364, pHC 364; L, linear; SC, supercoiled; CC, covalently closed circular; –, no added DNA. DNA-PK: +, DNA-PK added; –, no DNA-PK. Glucocorticoid receptor (GR): –, mock Sf9 extract; +, ammonium sulphate precipitate prepared from Sf9 cells expressing glucocorticoid receptor²⁷. **c**, Phosphorylation of GST-Oct-1-POU²⁸ (amino acids 280–439) by DNA-PK. Labelling is as for **b**, except that for Oct-1: +, GST-Oct-1-POU added; –, mock extract added. METHODS. DNA-PK phosphorylation of glucocorticoid receptor and GST-Oct-1 (ref. 28) were performed essentially by standard protocol²¹ using purified DNA-PK²¹ and 10 ng plasmid DNA. Glucocorticoid receptor and GST-Oct-1 were expressed in Sf9 cells and bacteria, respectively, and prepared as previously described^{22,27,28}. GST-Oct-1 resolved as a doublet on SDS-PAGE gels before kinasing (data not shown). For glucocorticoid receptor assays, it was immunoprecipitated by glucocorticoid antibody 59 (Affinity BioReagents) before SDS-PAGE while GST-Oct-1 was recovered on glutathione Sepharose before electrophoresis. DNAs were linearized with HindIII and in some instances recircularized with T₄ DNA ligase. All plasmid DNAs were gel-purified before use in kinase assays. Purified recircularized plasmids were resistant to S1 nuclease, ExoIII and Bal31.



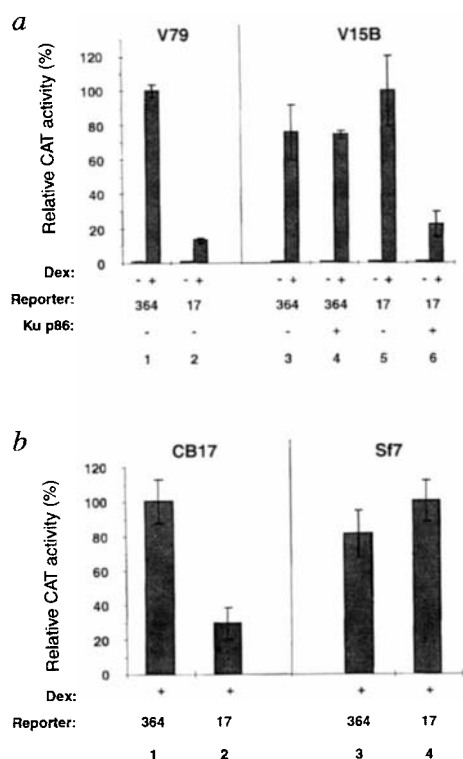


FIG. 4 Ku/DNA-PK repress MMTV expression through NRE1. *a*, Ku is required for NRE1 activity in hamster fibroblasts. Each lane shows CAT activity derived from transcription from the MMTV promoter plus (+, grey bars) and minus (-, black bars) glucocorticoid treatment. Reporter plasmids pHC17 (17) and pHC364 (364), which contain and lack NRE1, respectively, are shown in Fig. 3a. In V79 Chinese hamster fibroblasts, NRE1 repressed dexamethasone-induced MMTV transcription 9–10-fold (lanes 1 and 2). However, NRE1 did not mediate repression in p86⁻ XR-V15B cells⁶ (V15B, lanes 3–5), unless p86 Ku was re-expressed from a transfected cDNA (lane 6). Similar results were obtained in CHO K1 cells and the radiation-sensitive xrs5 derivative cell line that is also deficient in p86 Ku⁷ (data not shown). *b*, NRE1 fails to repress MMTV transcription in cells lacking DNA-PK. CAT activity derived from plasmids pHC17 and pHC364 in transfected wild-type CB-17 cells (lanes 1 and 2) and DNA-PK⁻ Sf7 SCID mouse fibroblast cells^{10,24} (lanes 3 and 4) treated with glucocorticoid hormone. In these cells, hormone-independent MMTV CAT activity was below measurable levels.

METHODS. Chinese hamster fibroblast cell lines V79 and XR-V15B (2×10^6 cells per plate) were transfected with MMTV/CAT plasmids pHC17 and pHC364 (ref. 14) ($4 \mu\text{g}$ per plate) and a rat glucocorticoid receptor expression vector ($2 \mu\text{g}$ per plate), using DEAE-dextran. Murine fibroblast cell lines CB-17 and Sf7 (5×10^5 cells per plate) were transfected with 100 ng reporter plasmid and $2 \mu\text{g}$ rat glucocorticoid receptor expression vector (each per plate) using lipofectamine (Gibco). p86 Ku was re-expressed from pBJ6-Ku80 (ref. 6) ($1.75 \mu\text{g}$ per plate) as indicated. Then, 16 h after transfection, cells were treated with 2×10^{-7} M dexamethasone for 48 h. CAT activity was corrected for β -galactosidase activity produced from a co-transfected RSV- β -gal plasmid ($2 \mu\text{g}$ per plate) and expressed as a percentage of maximum activity for each cell line. Activities are expressed as the means from 3–5 experiments. Error bars indicate 1 s.d.

modification to the subset of nuclear factors bound adjacent to NRE1-like sequences has the potential to create nuclear factor subpopulations with distinctive regulatory potentials. □

Received 14 December 1995; accepted 25 January 1996.

- Kuhn, A., Gottlieb, T. M., Jackson, S. P. & Grummt, I. *Genes Dev.* **9**, 193–203 (1995).
- Finnie, N., Gottlieb, T., Hartley, K. & Jackson, S. P. *Biochem. Soc. Trans.* **21**, 930–935 (1993).
- Genersch, E. et al. *EMBO J.* **14**, 791–800 (1995).
- Anderson, C. W. & Lees-Miller, S. P. *CRC Rev. Gene Express.* **2**, 283–314 (1992).
- Anderson, C. W. *Trends biochem. Sci.* **18**, 433–435 (1993).

- Smider, V., Rathmell, W. K., Lieber, M. R. & Chu, G. *Science* **266**, 288–292 (1994).
- Taccioli, G. et al. *Science* **265**, 1442–1445 (1994).
- Lees-Miller, S. P. et al. *Science* **267**, 1183–1185 (1995).
- Kirchgesner, C. U. et al. *Science* **267**, 1178–1183 (1995).
- Blunt, T. et al. *Cell* **80**, 813–823 (1995).
- Gottlieb, T. M. & Jackson, S. P. *Cell* **72**, 131–142 (1993).
- Pailard, S. & Strauss, F. *Nucleic Acids Res.* **19**, 5619–5624 (1991).
- Giffin, W., Torrance, H., Saffran, H., Macleod, H. L. & Haché, R. J. G. *J. Biol. Chem.* **269**, 1449–1459 (1994).
- Mink, S., Ponta, H. & Cato, A. C. B. *Nucleic Acids Res.* **18**, 2017–2024 (1990).
- Lee, J. W., Moffitt, P. G., Morley, K. L. & Peterson, D. O. *J. Biol. Chem.* **266**, 24101–24108 (1991).
- Yanagawa, S. I. et al. *J. Virol.* **67**, 112–118 (1993).
- Yaneva, M., Wen, J., Ayala, A. & Cook, R. J. *Biol. Chem.* **264**, 13407–13411 (1989).
- Falzon, M., Fewell, J. W. & Kuff, E. L. J. *Biol. Chem.* **268**, 10546–10552 (1993).
- Payvar, F. et al. *Cell* **35**, 381–392 (1983).
- Chalepakidis, G. et al. *Cell* **53**, 371–382 (1988).
- Lees-Miller, S. P., Sakaguchi, K., Ullrich, S. J., Appella, E. & Anderson, C. W. *Molec. cell. Biol.* **12**, 5041–5049 (1992).
- Schena, M., Freedman, L. P. & Yamamoto, K. R. *Genes Dev.* **3**, 1590–1601 (1989).
- May, G., Sutton, C. & Gould, H. J. *Biol. Chem.* **266**, 3052–3059 (1991).
- Rathmell, W. K. & Chu, G. *Molec. cell. Biol.* **14**, 4741–4748 (1994).
- Reeves, W. H. *Rheum. Dis. Clin. N.A.* **18**, 391–414 (1992).
- Ono, M., Tucker, P. W. & Capra, J. D. *Nucleic Acids Res.* **22**, 3918–3924 (1994).
- Srinivasan, G. & Thompson, E. B. *Molec. Endocr.* **4**, 209–216 (1990).
- Lai, J. S., Cleary, M. A. & Herr, W. *Genes Dev.* **6**, 2058–2065 (1992).

ACKNOWLEDGEMENTS. We thank G. Chu, M. Zdzienicka, P. Jeggo and G. Whitmore for cell lines; M. Connelly and C. Anderson for a sample of purified holo-DNA-PK; B. Thompson for baculovirus-expressed glucocorticoid receptor; F. Capra for baculovirus stocks of Ku; G. Chu, K. Yamamoto, W. Herr and A. Cato for plasmid DNAs; W. Reeves for Ku antibodies; and our colleagues in the Endocrine Research Group, C. Anderson and M. Beato for critical reading of this manuscript. This work was supported by the Medical Research Council of Canada. R.J.G.H. is a Scholar of the Medical Research Council of Canada and the Cancer Research Society Inc. H.T. and D.R. hold studentships from the Natural Sciences and Engineering Research Council of Canada.

Kinetic trapping of oxygen in cell respiration

Michael I. Verkhovsky, Joel E. Morgan, Anne Puustinen & Mårten Wikström

Helsinki Bioenergetics Group, Department of Medical Chemistry, Institute of Biomedical Sciences, and Biocentrum Helsinki, PB 8, 00014 University of Helsinki, Finland

CELL respiration in eukaryotes is catalysed by the mitochondrial enzyme cytochrome *c* oxidase. In bacteria there are many variants of this enzyme, all of which have a binuclear haem iron–copper centre at which O_2 reduction occurs, and a low-spin haem, which serves as the immediate electron donor to this centre¹. It is essential that the components of the cell respiratory system have a high affinity for oxygen because of the low concentrations of dissolved O_2 in the tissues; however, the binding of O_2 to the respiratory haem–copper oxidases is very weak^{2,3}. This paradox has been attributed to kinetic trapping during fast reactions of O_2 bound within the enzyme's binuclear haem iron–copper centre². Our earlier work³ indicated that electron transfer from the low-spin haem to the oxygen-bound binuclear centre may be necessary for such kinetic oxygen trapping. Here we show that a specific decrease of this haem–haem electron transfer rate in the respiratory haem–copper oxidase from *Escherichia coli* leads to a corresponding decrease in the enzyme's operational steady-state affinity for O_2 . This demonstrates directly that fast electron transfer between the haem groups is a key process in achieving the high affinity for oxygen in cell respiration.

At low temperatures, O_2 first binds to the enzyme reversibly to yield a haem Fe– O_2 'oxy' compound². Subsequent reactions within the oxygen-bound haem iron–copper centre have been proposed to trap the oxygen molecule kinetically². In this model, the operational K_M for O_2 (the concentration yielding half-maximal steady-state turnover) is a function of the ratio between the rate of steady-state turnover and the rate by which O_2 is trapped^{4,5}. We later confirmed that O_2 binding is weak and reversible at room temperature, but showed that fast electron transfer from the low-spin haem to the oxygen-bound binuclear

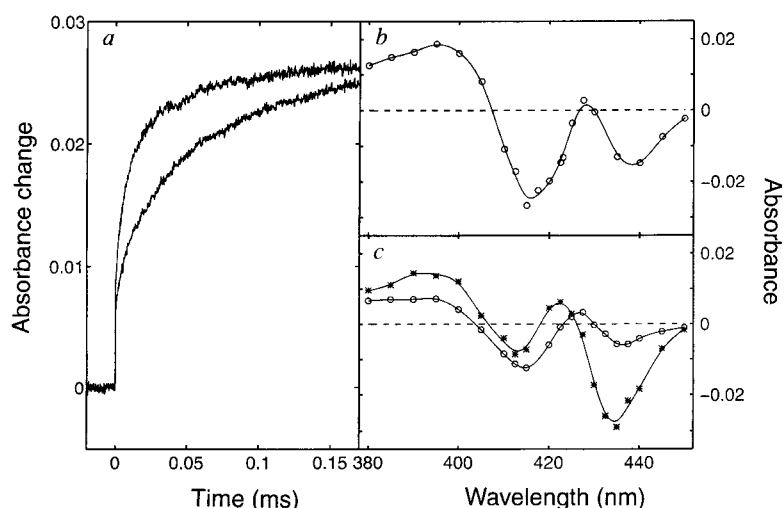
site is the step that commits the enzyme to catalysing further oxygen reactions³. Hence this haem-haem electron transfer may be essential for oxygen trapping to occur.

To test this idea, we took advantage of a natural variation in the haem-copper oxidase of *E. coli*. This enzyme incorporates either haem B or haem O (ref. 6) in the low-spin haem site, whereas the oxygen-reactive site always contains haem O (ref. 7). The respective enzyme forms are called cytochromes *bo*₃ and *oo*₃, signifying that they have different low-spin haems but the same oxygen-reactive haem. We found earlier that haem-haem electron transfer is several times faster in the *bo*₃ enzyme than it is in *oo*₃ (ref. 8). Here we use two enzyme preparations, one containing cytochrome *bo*₃ and the other *bo*₃ and *oo*₃ in a 30/70 ratio, to resolve the haem-haem electron transfer processes in the two types of enzyme and to determine their spectra and rates of reaction (Fig. 1). Electron transfer time constants of 7–9 μ s and 52 μ s were found for the *bo*₃ and *oo*₃ enzymes, respectively.

FIG. 1 Electron transfer in cytochromes *bo*₃ and *oo*₃. **a**, Kinetics of absorbance changes at 400 nm in response to laser flash illumination of CO mixed-valence enzyme. Top curve, cytochrome *bo*₃; bottom curve, 30/70 mixture of cytochromes *bo*₃ and *oo*₃. Electron redistribution follows photolysis of CO from the enzyme in which the CO-reactive haem iron-copper site is reduced, but the low-spin haem (*b* or *o*) is oxidized. The immediate jump in absorbance is related to photolysis of the Fe-CO bond. The subsequent changes in the absorbance can be assigned to haem-haem electron transfer events for both types of enzyme. **b**, Kinetic spectrum of the major component during microsecond absorbance changes for cytochrome *bo*₃. A global kinetic fit finds a single fast phase ($\tau \approx 7 \mu$ s), the spectrum of which indicates that it is due to electron transfer from haem *o*₃ to haem *b*. The broad, high-spin ferrous haem *o*₃ absorbance disappears and is replaced by the narrower peak of ferrous low-spin haem *b* at 428 nm. **c**, Two kinetic phases emerge in the *bo*₃/*oo*₃ enzyme mixture. The smaller fast phase (circles; $\tau \approx 9 \mu$ s) has the same spectrum as the *bo*₃ enzyme. The dominant slower phase (asterisks; $\tau \approx 52 \mu$ s) can be assigned to electron transfer from haem *o*₃ to haem *o* (that is, in the cytochrome *oo*₃ majority population). Here again, the broad high-spin haem *o*₃ signal disappears, but now the emerging peak has a maximum at 422 nm, indicating that low-spin haem *o* has been reduced.

METHODS. As a source of pure cytochrome *bo*₃, we used a strain of *E. coli* (GO 105) bearing the plasmid pJRH1A, in which a C-terminal histidine tag had been genetically added to subunit II to aid enzyme isolation in a single step in a metal chelate affinity column (J. N. Rumbley and R. B. Gennis, unpublished results). The enzyme from strain RG 129 (bearing plasmid pMC31; ref. 12) is a mixed population of 70% cytochrome *oo*₃ and 30% cytochrome *bo*₃ (values determined by the pyridine haemochrome method^{7,13}). Experiments were carried out in 100 mM MOPS buffer, pH

Despite this 6–7-fold difference in the rate of haem-haem electron transfer, there is no significant difference in the maximum steady-state rates of enzyme turnover between the two preparations (Fig. 2), so turnover is not limited by haem-haem electron transfer. On the other hand, when the reduced enzyme reacts with O₂, the early O₂-concentration-dependent electron transfer event is also slower by a factor of nearly ten in the *oo*₃ enzyme compared with that in *bo*₃ (compare refs 3,9). Thus haem-haem electron transfer influences the rate at which the low-spin haem is initially oxidized in the oxygen reaction, even though the latter reaction is always much slower than electron transfer measured directly after CO photolysis in the absence of O₂ (refs 3,10,11). The fact that the rate of this reaction depends on both the electron transfer rate and on the O₂ concentration shows that the interaction of the enzyme with O₂ must be effectively reversible up to the point where the haem-haem electron transfer takes place³.



6.6, 0.02% n-dodecyl- β -D-maltoside. Enzyme concentrations were 3.13 μ M for *bo*₃ and 3.42 μ M for the *bo*₃/*oo*₃ mixed enzyme population. Transient absorption measurements were made using a dual-beam microsecond apparatus⁸. A camera-type xenon flash lamp was used as the probe light source. A frequency-doubled YAG laser (Quintel Brilliant) was used for photolysis. For data analysis, Graphic Interactive Management (GIM, Alexander Drachev), Matlab (The Mathworks, South Natick, Massachusetts), and a global kinetic fit programme (independent executable programme run under a Matlab front-end) was used, which assumes that kinetic behaviour at all wavelengths can be described in terms of a small number (2 to 4) of exponential components¹⁴.

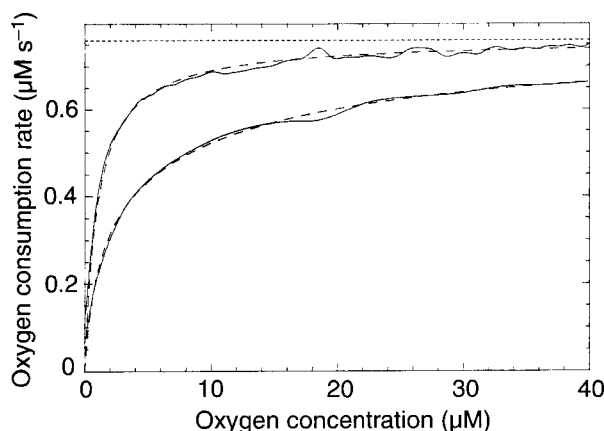


FIG. 2 Dependence of the steady-state rate of respiration on the oxygen concentration. Top curve, cytochrome *bo*₃; lower curve, mixed (30/70) population of cytochromes *bo*₃ and *oo*₃. Measurements were made in 100 mM MOPS buffer, pH 7.0, 0.05% n-dodecyl- β -D-maltoside, 0.1 mM ubiquinone-1, 5 mM dithiothreitol, 49 nM enzyme. Dashed lines show the results of a nonlinear regression fit with the following parameters: K_M (oxygen concentration at half-maximal rate), 1 μ M for cytochrome *bo*₃ and 6 μ M for cytochrome *oo*₃; maximal rate (dotted line), 0.74 μ M O₂ per second.

Assuming, then, that kinetic trapping of O_2 to the enzyme is due primarily to haem-haem electron transfer, the apparent K_M can be approximated as (see text and refs 4,5)

$$K_{M,app} \approx \frac{k_{in}K_{eq}}{k_{et}} \quad (1)$$

where k_{in} is the rate constant for the largely rate-determining electron input into the enzyme, k_{et} is the rate constant of haem-haem electron transfer, and K_{eq} is the equilibrium constant of O_2 binding. Equation (1) predicts that a 6–7-fold decrease in k_{et} , such as would occur in the *E. coli* enzyme if haem O instead of haem B was at the low-spin haem site (Fig. 1), should increase $K_{M,app}$ by the same factor. We therefore tested this prediction (Fig. 2). In the cytochrome bo_3 preparation (top trace), the dependence of the

respiratory rate on O_2 concentration is monophasic, with a $K_{M,app}$ of 1 μM . In contrast, in the bo_3/oo_3 mixture the dependence is biphasic. Correcting for the 30% contribution of cytochrome bo_3 , we find that the $K_{M,app}$ of the oo_3 population is 6 μM . Thus the operational oxygen affinity is decreased by a factor very close to that by which haem-haem electron transfer is slowed down.

We conclude that fast haem-haem electron transfer is a key determinant of the enzyme's oxygen affinity. The strategy of cell respiration is thus not to invest energy in tight binding of O_2 , but instead to trap it kinetically by a redox process that requires fast haem-haem electron transfer. In this way, the high oxygen affinity that is essential for normal tissue function is realized simultaneously with the effective harnessing of energy released in cell respiration for the production of ATP. $\square$

Received 25 October 1995; accepted 24 January 1996.

1. Babcock, G. T. & Wikström, M. *Nature* **356**, 301–309 (1992).
2. Chance, B., Saronio, C. & Leigh, J. S. *J. biol. Chem.* **250**, 9226–9237 (1975).
3. Verkhovsky, M. I., Morgan, J. E. & Wikström, M. *Biochemistry* **33**, 3079–3086 (1994).
4. Chance, B. *J. gen. Physiol.* **49**, 163–188 (1965).
5. Petersen, L. C., Nicholls, P. & Degn, H. *Biochem. J.* **142**, 247–252 (1974).
6. Puustinen, A. & Wikström, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 6122–6126 (1991).
7. Puustinen, A. *et al. Biochemistry* **31**, 10363–10369 (1992).
8. Morgan, J. E., Verkhovsky, M. I., Puustinen, A. & Wikström, M. *Biochemistry* **32**, 11412–11418 (1993).
9. Svensson, M. & Nilsson, T. *Biochemistry* **32**, 5442–5447 (1993).

10. Hill, B. C. & Greenwood, C. *Biochem. J.* **218**, 913–921 (1984).
11. Oliveberg, M. & Malmström, B. G. *Biochemistry* **30**, 7053–7057 (1991).
12. Lemieux, L. J., Calhoun, M. W., Thomas, J. W., Ingledew, W. J. & Gennis, R. B. *J. biol. Chem.* **267**, 2105–2113 (1992).
13. Berry, E. A. & Trumpower, B. L. *Analyt. Biochem.* **161**, 1–15 (1987).
14. Provencer, S. W. & Vogel, R. H. in *Progress in Scientific Computing* Vol. 2 (eds Deuffhard, P. & Hairer, E.) 304–319 (Birkhäuser, Boston, 1983).

ACKNOWLEDGEMENTS. Jon N. Rumbley and Robert B. Gennis kindly provided the plasmid pJRHSA. This work was supported by the Academy of Finland, the University of Helsinki, and the Sigrid Jusélius Foundation.

KNOW YOUR COPY RIGHTS RESPECT OURS

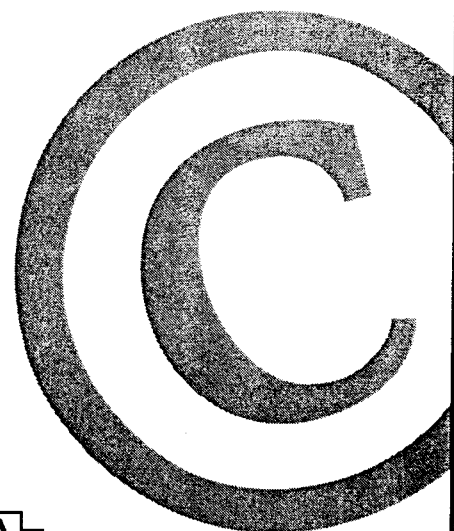
The publication you are reading is protected by copyright law. This means that the publisher could take you and your employer to court and claim heavy legal damages if you make unauthorised photocopies from these pages.

Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

The Copyright Licensing Agency (CLA) is an organisation which issues licences to bring photocopying within the law. It has designed licensing services to cover all kinds of special needs in business, education, and government.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931
Fax: 0171 436 3986